POLYSACCHARIDE UTILIZATION BY THE HUMAN COLONIC BACTERIUM, *BACTEROIDES INTESTINALIS* DSM 17393

BY

GABRIEL VASCONCELOS PEREIRA

DISSERTATION

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Isaac Cann, Chair
Professor Roderick Mackie
Assistant Professor Jason Ridlon
Professor Cari Vanderpool

ABSTRACT

The human gastrointestinal microbiome has co-evolved with the host for an extended period, generating an intrinsic metabolic network capable of modulating both the microbial community and host physiology. Changes in the gut environment shape the microbial community with one of the main driving characteristics being competition for nutrients. A key aspect for nutrient acquisition in the gut is the ability of the microbes to degrade and ferment dietary fiber. These chemically diverse polysaccharides are the main components of the plant cell wall and function as the primary energy source for the microbiota. The dietary polysaccharides are largely indigestible by the digestive enzymes of the host. The phylum Bacteroidetes, one of the main colonizers of the gastrointestinal tract (GIT), has evolved one of the largest arrays of carbohydrate-associated enzymes (CAZymes) responsible for degrading polysaccharides. Furthermore, these bacteria have highly organized gene clusters, known as polysaccharide utilization loci (PUL), encoding the enzymes required for the degradation and transport of the sugar components of dietary fiber. These PULs also possess an unusual regulatory mechanism found only in Bacteroidetes, similar to the canonical two-component system encoded by organisms in all three domains of life. The two-component system is involved in signal transduction, mediated by a histidine-kinase sensor and a response regulator. Interestingly, the Bacteroidetes encode this system in a single polypeptide that can be delineated into a sensor, a Y_Y_Y, a histidine-kinase, a histidine-ATPase, and response regulator modules, respectively.

Through our work, we aim to understand the degradation of esterified arabinoxylan by diverse *Bacteroides* spp. and to develop a rapid approach for identifying the potential

target of a vast number of uncharacterized PULs. Our findings demonstrate that some *Bacteroides* spp. highly express an esterase-enriched cluster during growth on esterified arabinoxylan compared to the mixture of its component monosaccharides xylose and arabinose. Biochemical analysis of the proteins encoded by the esterase-enriched cluster demonstrated diverse enzymatic activities and substrate specificities capable of working synergistically to fully depolymerize complex arabinoxylan substrates. Interestingly, the hypothetical domain encoded by BACINT_01040 demonstrated one of the most versatile feruloyl esterase activities, and the enzyme was able to cleave the ferulic acid side chain of every substrate tested in our studies. Moreover, we demonstrated that the bacteria harboring the esterase-enriched cluster do not metabolize the ferulic acid during growth on wheat bran. Thus, the plant phenolic compound accumulated in the spent medium. Furthermore, the accumulated ferulic acid was able to modulate the immune system and induce a Th1-type immune response and viral defenses in both gastrointestinal cell lines and mouse model.

Due to the vast number of uncharacterized PULs encoded in the microbiome, it is important to develop a rapid means to assign function to these gene clusters. Our current hypothesis relies on the observation of a potential endo-acting enzyme near the *susC*/*susD*-like genes. We hypothesized that these enzymes “scout” the environment for their associated target polysaccharides. In this work, we biochemically characterized the activity of two “scouting enzymes” on arabinoxylan and arabinan, thus showing that a PUL-associated hybrid two-component system (HTCS) sensor domain is able to bind and sense the products of its associated scouting enzyme. This hypothesis was further corroborated by transcriptomic analysis of *B. intestinalis* grown on each polysaccharide

compared to its monomeric components. Thus, the results demonstrated in this study may advance the understanding of polysaccharide utilization by Bacteroidetes species by rapidly assigning function to uncharacterized PULs. Further studies may also allow us to identify a polysaccharide-degrading signature by the Bacteroidetes species, which may be used for personalized dietary interventions and modulation of the gut community.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

The interaction between the gastrointestinal tract and the resident microbiota has been extensively studied over several decades. The major focus of these studies has been on understanding the effects of the microbiota in the human gut and its impact on host physiology, immune system, and nutrition. These interactions, which happen through different routes such as direct interaction of microbes and the epithelium lining, as well as through metabolites produced by fermentation or synthesized by the gut microbiota, have been fine-tuned through a parallel evolution between symbiotic organism and the host (Xu, Mahowald et al. 2007, Ley, Hamady et al. 2008). The host evolved to utilize short-chain fatty acids as an energy source as well as signaling molecules for growth and development, while the bacteria evolved to take advantage of nutrients that evade the host gastrointestinal system (Guarner and Malagelada 2003). Furthermore, this complex host/microbe interaction possesses a very dense bacterial population, reaching up to 10^{12} cells per gram of stool. The human gastrointestinal tract one of the most microbially colonized environments (Savage 1977). Although the human gastrointestinal tract (GIT) is a diverse environment, it is mainly dominated by two different phyla, the Bacteroidetes and the Firmicutes (Turnbaugh, Backhed et al. 2008, Turnbaugh, Hamady et al. 2009). Recent research has demonstrated large variances in the ratio of Bacteroidetes:Firmicutes, mainly attributed to host diet and genetic background. There are clear benefits of a stable microbial community to the host and shifts in the members, as well as the metabolic functions in the GIT have been shown to disrupt the microbial

community leading to colonization by opportunistic pathogens as well as metabolic disorders which are detrimental to the host (Louis, Hold et al. 2014).

Within the functions employed by bacteria that inhabit the human GIT, degradation of dietary fiber is one of the key aspects for the ability to colonize and thrive in this competitive environment (El Kaoutari, Armougom et al. 2013). Dietary fiber and host glycans, the main polysaccharide source for bacterial fermentation, are composed of complex glycosidic linkages that require a complex enzymatic array to release the fermentable monomers. The Bacteroidetes phylum in the human colon possesses the largest array of glycoside hydrolases and polysaccharide lyases, responsible for cleaving glycosidic linkages. These bacteria encode gene clusters designated polysaccharide utilization loci (PUL) and are known to target specific polysaccharides (Martens, Chiang et al. 2008). These PULs are usually composed of transporters (SusC-like), carbohydrate binding proteins (SusD-like), glycoside hydrolases, carbohydrate esterases, and regulators (hybrid two-component system, SusR-like, sigma and anti-sigma factors). Some Bacteroidetes species, for example *Bacteroides intestinalis*, encodes about 68 different PULs that likely targets different polysaccharides. Therefore, the PULs allows this phylum to degrade a wide range of host and dietary glycans (El Kaoutari, Armougom et al. 2013).

Currently, there are roughly 5,000 PULs deposited in the Polysaccharide Utilization Loci Database (http://www.cazy.org/PULDB/) from close to 900 different Bacteroidetes species. Each of these PULs contains several different genes, mostly with unassigned function as well as unassigned targets. The main regulator of these PULs, the hybrid two-component system, is a polypeptide that seems to act as the canonical two-component

system encoded by several different bacteria (Hoch and Silhavy 1995). The two-component system utilizes a sensor protein in the cytoplasmic region combined with a cytoplasmic phosphorylating cascade of the response regulator. In the HTCS, this mechanism is encoded by a single polypeptide and understanding the substrate specificity of the sensor domain in the cytoplasm can lead to assignment of function to a wide range of PULs.

Plant-derived polysaccharides are among the largest repositories of carbon source and are an integral part of human diets. The large variety of plant cell wall structures highlights the complexity of linkages, which hinders the host's ability to extract nutrients from most polysaccharides (Dodd and Cann 2009). With the exception of starch, most of the dietary fiber consumed by the host evades degradation and thus reach the colon. One of the main components of plant cell wall hemicellulose, xylan, is composed of a β-1,4 linked xylose backbone with several different decorating moieties such as arabinose, glucuronic acid, and ferulic acid. Arabinoxylan is a polysaccharide present in different human diet-related cereals such as wheat, oats, rye, and corn (Izydorczyk and Biliaderis 1995). Therefore, arabinoxylan degradation by the resident gut microbiota is key to host homeostasis and physiology. Recent studies have demonstrated interest in natural phenolic compounds as a means to decrease inflammation in the host, which may be the cause of several chronic diseases such as diabetes, inflammatory bowel disease (IBD) and others (Yasuko, Tomohiro et al. 1984, Serrano, Palacios et al. 1998). Arabinoxylan substrates are a rich source of different phenolic compounds, mainly ferulic acid. Thus understanding the degradation of this polysaccharide can help manipulate of host health.

1.2 Human lower gastrointestinal tract

The co-evolution of humans with the trillions of bacteria that dwell in our bodies generated a complex, yet stable, metabolic network that affects host physiology and health. Moreover, humans demonstrate only small differences in their genetic composition, yet they exhibit large phenotypic variations. Distinctively, the microbiome composition presents a large variability even between siblings (Turnbaugh, Hamady et al. 2009), creating a great challenge in the microbiome research of a healthy "core microbiome". Several studies point to a genetic "functional core", in which different organisms are capable of providing similar metabolic functions that influence the host-microbiome interaction (Lloyd-Price, Abu-Ali et al. 2016). At present, these functions are usually associated with dietary fiber metabolism, production of short chain fatty acids (SCFA), secondary metabolites, immune maturation of the host, and vitamin production. Dysbiosis of this close interaction between microbes and the host is associated with several diseases, such as diabetes, inflammatory bowel disease, allergies, cancer, and other conditions (Guinane and Cotter 2013, Karlsson, Tremaroli et al. 2013, Louis, Hold et al. 2014, West, Jenmalm et al. 2015). Similar to pathogenicity, these perturbations of the microbiome equilibrium can exacerbate or induce disease states. Bacteria perform a wide variety of biological activities that are absent in the human host (Gill, Pop et al. 2006, Qin, Li et al. 2010, Hooper, Littman et al. 2012, Human Microbiome Project 2012, Karlsson, Tremaroli et al. 2013). These microbe derived biological activities are executed by a large, reportedly, non-redundant gene set that contains more than 3.3 million microbial genes, approximately 150-fold more than humans genes (Qin, Li et al. 2010). Even though, bacteria possess this vast genetic capability, roughly 75% of the genes

encoded by the human microbiome community either do not have attributed function or are poorly characterized (Qin, Li et al. 2010).

Within the human microbiome, the ecosystem of the colon has been studied intensively, due to the vast microbial biomass that surpasses that of any other body site. Members of the microbes present in the lower GIT are capable of fermenting dietary fibers that evade host digestion, due to the recalcitrant structure of plant polysaccharides (Flint, Scott et al. 2012). The fermentation of plant materials such as cellulose, hemicellulose, and starch in the GIT produces three major SCFAs, i.e. acetic, butyric, and propionic acids (Bergman 1990), which serve as nutrients for colonic epithelial cells (Topping and Clifton 2001). These fermentation products account for approximately 3-10% of the human caloric intake (Bergman 1990). Furthermore, the large diversity of the microbiota within the lower GIT generates an evolutionary pressure and has a large impact on microbial diversity. Despite this large diversity within the GIT environment, human colonic microbiomes are dominated by Bacteroidetes and Firmicutes (Qin, Li et al. 2010, Human Microbiome Project 2012, Faith, Guruge et al. 2013), although the ratio between these two bacterial phyla varies greatly among individuals (Human Microbiome Project 2012).

***Bacteroides* in the human gut.** Compared to the human microbiome from different sites, the colon contains the largest and most diverse microbial community, harboring about 10^{11} organisms per gram of stool, of which it is estimated that 25% belong to the *Bacteroides* genus (Salyers 1984). These bacteria are gram-negative rods, anaerobic, bile acid-resistant, and non-spore-forming (Wexler 2007). Since its discovery, the obligate anaerobic *Bacteroides* genus has received considerable attention in human gut microbiology [18-20], due to their robust adaption to the gut environment, stable

colonization, and health benefits to the host (Bjursell, Martens et al. 2006). The host GIT presents a harsh environmental challenge to invading microorganisms that colonize the colon. This includes a strong pH gradient, from pH 2 at the stomach to a neutral pH 6-7 at the colon (Nugent 2001), as well as a three-dimensional oxygen gradient from the upper to the lower GIT, and that also varies from epithelium to lumen (Tegtmeier, Thompson et al. 2015). It is important to note that several *Bacteroides* encode the enzyme catalase that could allow the reduction of intracellular hydrogen peroxide levels as well as the colon environment (Rocha, Selby et al. 1996).

As nutrients flow to the colon, it is important to note that the readily accessible sugars would have already been absorbed or consumed by the host or other microbes living in the upper GIT. Thus, *Bacteroides* have adapted to degrade and consume the remaining complex plant polysaccharides and host glycans (Flint, Scott et al. 2012, Zhang, Chekan et al. 2014, Wang, Pereira et al. 2016). The mechanism utilized to break down complex polysaccharides, currently unique to the phylum Bacteroidetes, is composed of gene clusters termed polysaccharide utilization loci (PULs). The first identified PUL, the *Bacteroides thetaiotaomicron* starch utilization system or SUS, discovered by the Salyers's laboratory (Kotarski and Salyers 1984, Reeves, DElia et al. 1996), was composed of an eight-gene cluster that targets starch. The gene products of the SUS are located in the periplasm and outer membrane, and work synergistically to bind, degrade substrates into smaller fragments, regulate gene transcription, transport products into the periplasmic region, and further degrade products into simple sugars for fermentation in the cytoplasm.

Based on the starch utilization system and its composition, PULs were classified as gene clusters that contain the homologs of SusC/SusD pair (Terrapon, Lombard et al. 2015). The advances in bioinformatics techniques as well as next generation sequencing, allowed the prediction of more than 3000 PULs encoded in 67 Bacteroidetes from the human gut (Terrapon, Lombard et al. 2015). These advances in sequencing resulted in the complete genome sequence of several members of the *Bacteroides* genus, including *Bacteroides thetaiotaomicron* (Xu, Bjursell et al. 2003), *Bacteroides fragilis* (Kuwahara, Yamashita et al. 2004, Cerdeno-Tarraga, Patrick et al. 2005), *Bacteroides ovatus* (Wu, McNulty et al. 2015), *Bacteroides cellulosilyticus* (Wu, McNulty et al. 2015), and *Bacteroides intestinalis* [ABJL00000000.2-].

Polysaccharide utilization by genus *Bacteroides.* The carbohydrate utilization by microbes is directly linked to their ability to persist in complex and competitive environments such as the human colon. Thus, polysaccharide utilization loci in the phylum Bacteroidetes embodies a great advantage in adaptability towards diverse polysaccharides degradation. These systems are organized gene clusters, that are tightly regulated and that encode enzymes required for depolymerization of several different polysaccharides. Since most of the nutrients are absorbed or consumed before reaching the colon, mainly complex oligo- and polysaccharides reach the lower GIT. These complex carbohydrates are composed of diverse sugar units, linkages, and decorating moieties. Therefore, to completely breakdown these carbohydrates into fermentable sugars, the *Bacteroides* utilize an equally diversified array of enzymes (Figure 1.1). Recent work by Ndeh et al, showed that *B. thetaiotaomicron* is able to cleave 20 different

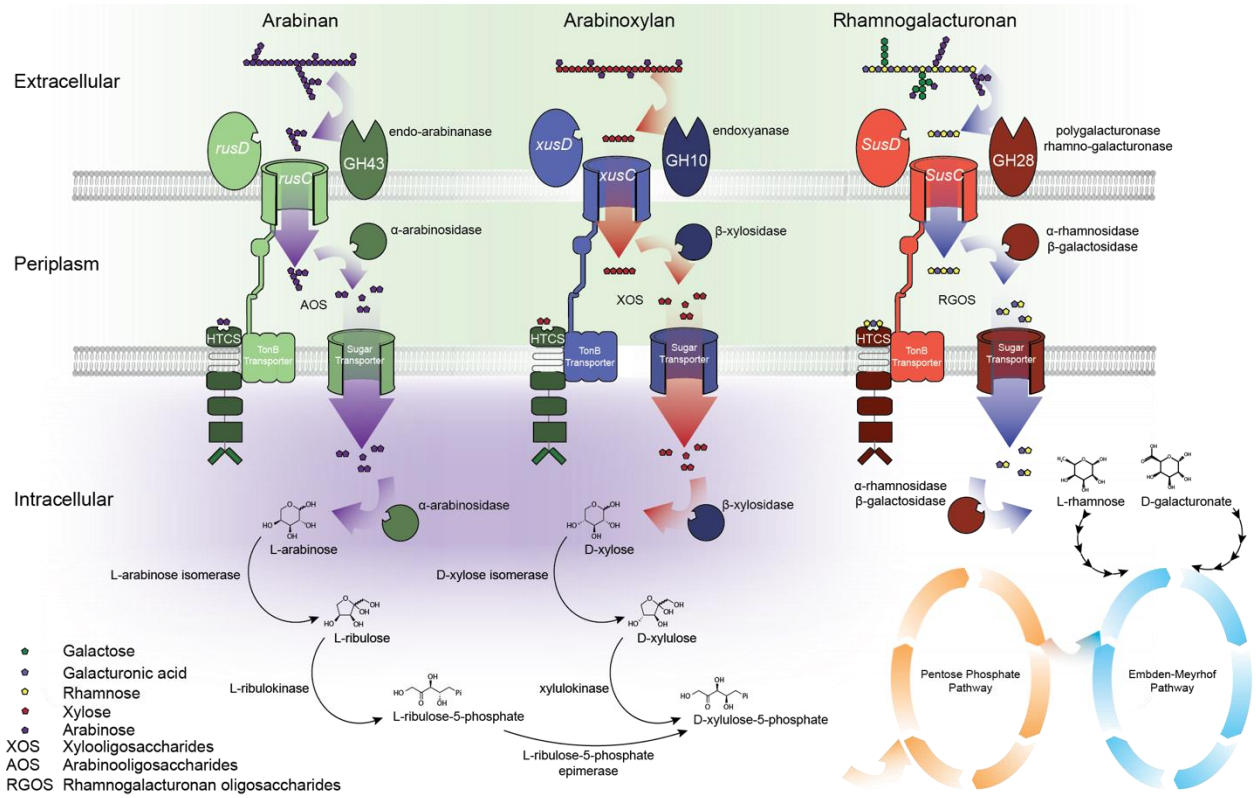


Figure 1.1 - **Polysaccharide utilization loci model.** A cell surface enzyme that cleaves a unique polysaccharide present in the environment of the cell is shown for each PUL. The enzymes are an endo-arabinanase, an endoxylanase, and a polygalacturonase and rhamnogalacturonase, respectively, for the arabinan, arabinoxylan, and RGI and HG PULs. The current model is that in the absence of the polysaccharide, the HTH of the HTCS is bound to the DNA controlling the expression of the genes in the cluster, and thus blocks production of enzymes in the cluster. In the presence of polysaccharide in the environment, the cell-surface-associated enzyme cleaves the polysaccharide, generating oligosaccharides that are transported into the periplasmic space. The sensor module of the HTCS senses the shorter chain products through binding and releases the HTH module from the DNA, triggering expression of enzymes in the cluster. Enhanced gene expression in the cluster leads to extensive hydrolysis of the polysaccharide to monomeric sugars that are then fermented by the bacterium through different metabolic pathways, as shown.

linkages from the plant pectic polysaccharide rhamnogalacturonan-II (Ndeh, Rogowski et al. 2017).

The PULs are vital for the persistence and colonization of the human lower gut, comprising up to 30% of the genome in certain *Bacteroides* species. Currently, the Carbohydrate-Active Enzymes (CAZyme) database contains 135 families of Glycoside Hydrolases (GHs), 24 families of Polysaccharide Lyases (PLs), 16 families of Carbohydrate Esterases (CEs), and 101 families of Glycosyltransferases (GTs) (Grondin, Tamura et al. 2017). Classified as a group of enzymes that catalyzes the cleavage of glycosidic linkages of an O-glycoside bond, glycoside hydrolases are found in almost every living organism. The array of GHs in a microorganism is directly correlated with their environment and niche. These genes are prone to duplications and horizontal gene transfer, varying even between strains. Within the glycoside hydrolases, small changes in structure are able to shift the substrate specificity of the enzyme, as well as evolutionarily related enzymes which may bind different substrates. The two known catalytic mechanisms of GHs result in a retaining or inversion of the carbohydrate anomeric configuration. During the retaining cleavage, a double displacement reaction takes place in two transition states. Initially an ester is formed among the enzyme nucleophile (asparagine or glutamic acid) and the C1 of the targeted sugar. During the second stage, the ester bond is hydrolyzed, releasing the substrate in the initial anomeric configuration (Figure 1.2A) (Jordan and Wagschal 2010). In contrast, the inversion mechanism occurs through a one-step single-displacement reaction, releasing the substrate with a different anomeric configuration (Jordan and Wagschal 2010).

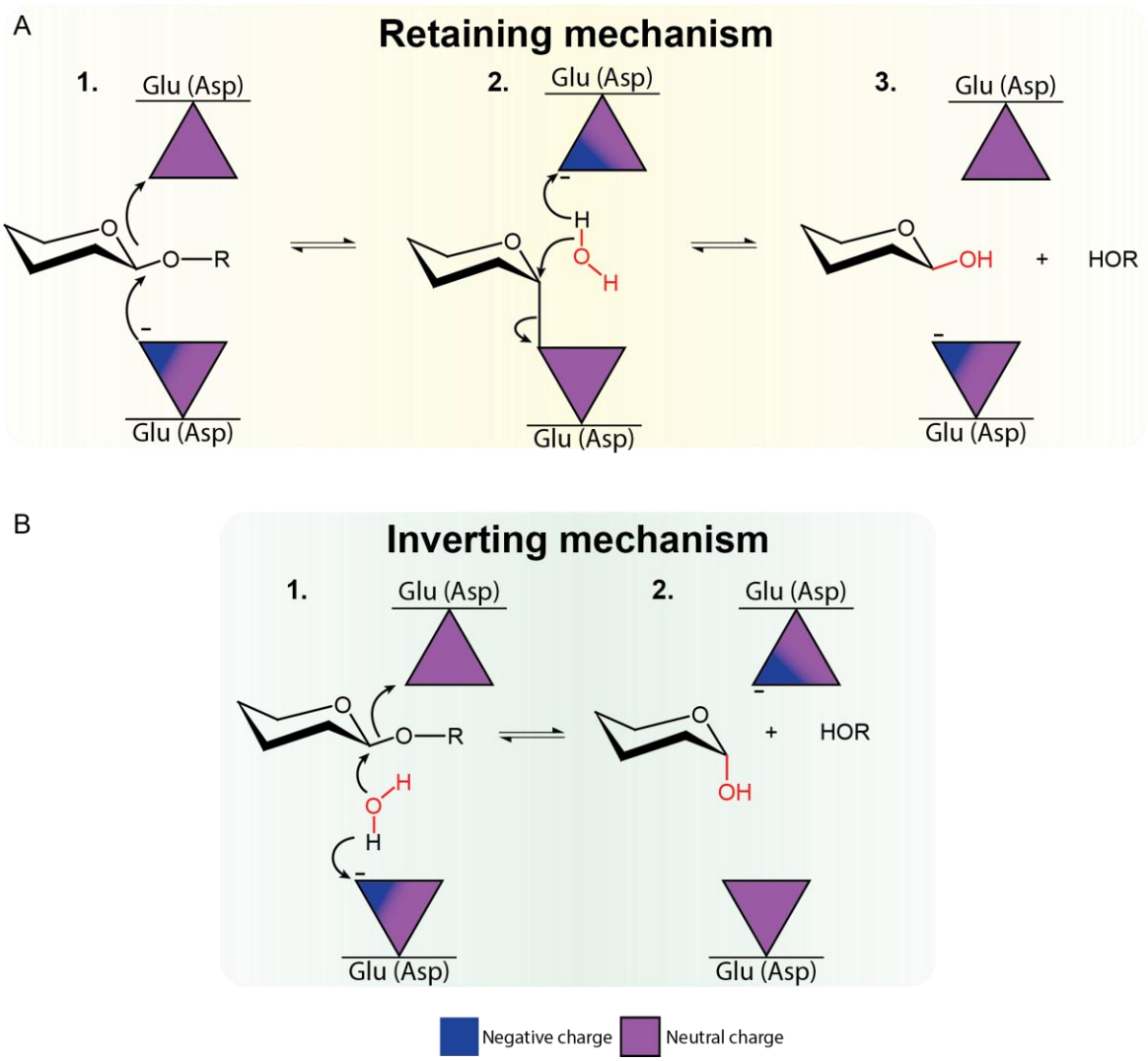


Figure 1.2 **Glycoside hydrolase mechanisms.** (A) The retaining mechanism start with the (**1.**) displacement of the attached sugar (R) by the deprotonated carboxylic group, generating (**2.**) a covalent transition state of sugar-enzyme. Posteriorly, the activated water molecule act as a nucleophile, displacing the covalent sugar-enzyme bond, (**3.**) maintaining the sugar sterochemical configuration. (B) In the inverting mechanism, the activated (**1.**) water displaces the attached sugar, generating an inversion of the (**2.**) stereochemical configuration.

Consequently, these enzymes are key for degradation of polysaccharides and are vital components of the polysaccharide utilization loci encoded by *Bacteroides*, in which these bacteria encode an average of 2.5-4 GHs per PUL (Terrapon, Lombard et al. 2015).

Carbohydrates Esterases are enzymes that hydrolyze ester bond linkages between carbohydrates and its moieties, catalyzing de-O or de-N-acylation (Cantarel, Coutinho et al. 2009). Currently, there are 15 different carbohydrate esterase families in the Carbohydrate-Active Enzyme database. There is a vast application for the carbohydrate esterases, both physiologically and industrially. These enzymes catalyze the removal of ester-linked decorations to carbohydrates, in which these decorations hinder the cleavage activity of glycoside hydrolases towards the sugar linkages in the carbohydrates. Thus, removal of these acylated moieties of the polysaccharides increases the rate of depolymerization by synergistically working with the GHs. Within this class of enzymes, there are a large variety of substrate specificities and activities such as acetylxylan esterase, feruloyl esterase, chitin deacetylases, and pectin acetyl esterases. Feruloyl esterases are enzymes that cleave the hydroxycinnamic acids esterified to arabinoxylans and some pectins present in plant cellular wall (de Vries, vanKuyk et al. 2002). This enzymatic reaction is key for organisms to fully degrade these polysaccharides into fermentable sugars. Acetyl xylan esterases remove the acetyl group esterified to xylan polysaccharide, allowing better accessibility of glycoside hydrolases such as endoxylanases to their substrates. There is a wide range of these enzymes in bacteria and fungi, and they are highly inducible by different substrates containing these esterified moieties (Kabel, Yeoman et al. 2011).

The mechanism employed by carbohydrate esterases, especially feruloyl and acetyl xylan esterases, is similar to the catalytic triad mechanism by serine proteases (Kraut 1977). This enzymes utilize the amino acids Ser-His-Asp(Glu) in a catalytic triad in two steps to cleave the ester linkages in the polysaccharide backbone. The initial phase, acylation of the serine, occurs through an electron relay from the deprotonated carboxylic group from Asp(Glu) to the His and Ser, allowing the acylation of the serine and displacement of the sugar backbone. In the second step, the deacylation of the serine, occurs through an activated water molecule that attacks the ester linkage of the acyl-serine, combined with an electron relay through the His to Asp(Glu) and returning to the original state (Figure 1.3).

Another important aspect of the polysaccharide utilization loci is the ability to orchestrate import of oligosaccharides derived from polysaccharide degradation that occurs in the extracellular region into the periplasmic region. This oligosaccharide import is the function of a protein complex named SusC/SusD pair in reference to the starch utilization system found by Salyers et al (Kotarski and Salyers 1984). This protein complex is found in many *Bacteroides* spp., with some examples of *Bacteroides* encoding over 100 SusC/SusD pairs that exhibit a wide variety of substrate specificity. Recent studies demonstrate that SusC is a homodimer TonB-dependent transporter, in which SusD interacts by capping the two β-barrel protomer. SusD is a extracellular protein that contains a substrate binding domain that without substrate, moves away from the SusC transporter channel to expose the substrate-binding site in a hinge-like motion, while during substrate binding it closes the lid by the stabilized interaction of SusC/SusD/substrate (Figure 1.4) (Glenwright, Pothula et al. 2017). Moreover, SusD is

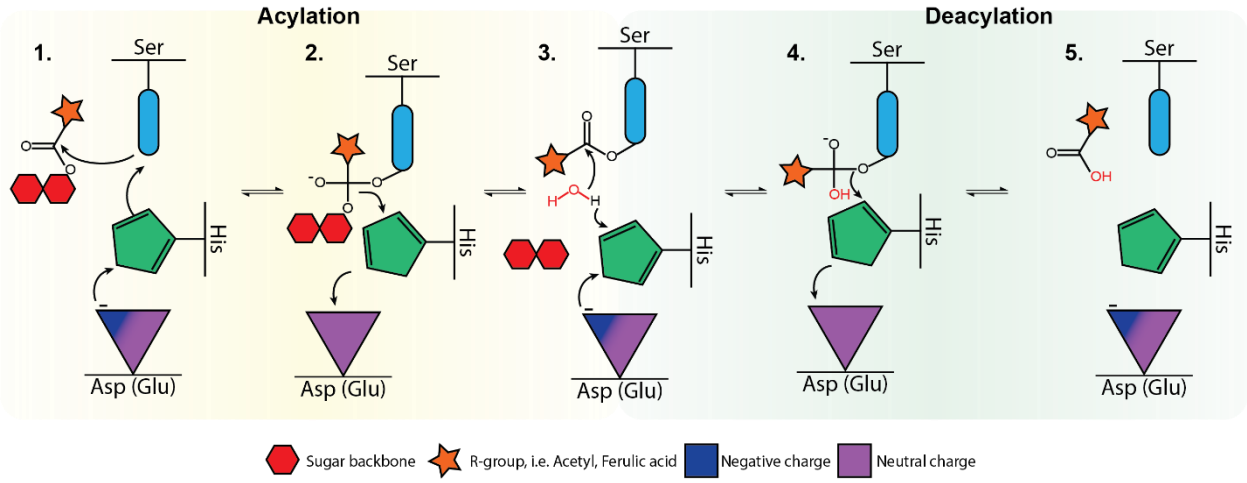


Figure 1.3 **Schematic of esterase mechanism employing the catalytic triad composed of Ser-His-Asp(Glu).** Initially, (**1.**) an electron relay from the deprotonated carboxylic group is transferred from the Asp(Glu) to His and Ser, (**2.**) allowing the acylation of the enzyme serine and displacement of the sugar backbone (red). An (**3.**) activated water attacks the ester linkage of the acyl-serine, (**4.**) combined with an electron relay that deprotonate the Asp(Glu) carboxylic acid results in the deacylation of the serine and return to the (**5.**) original status.

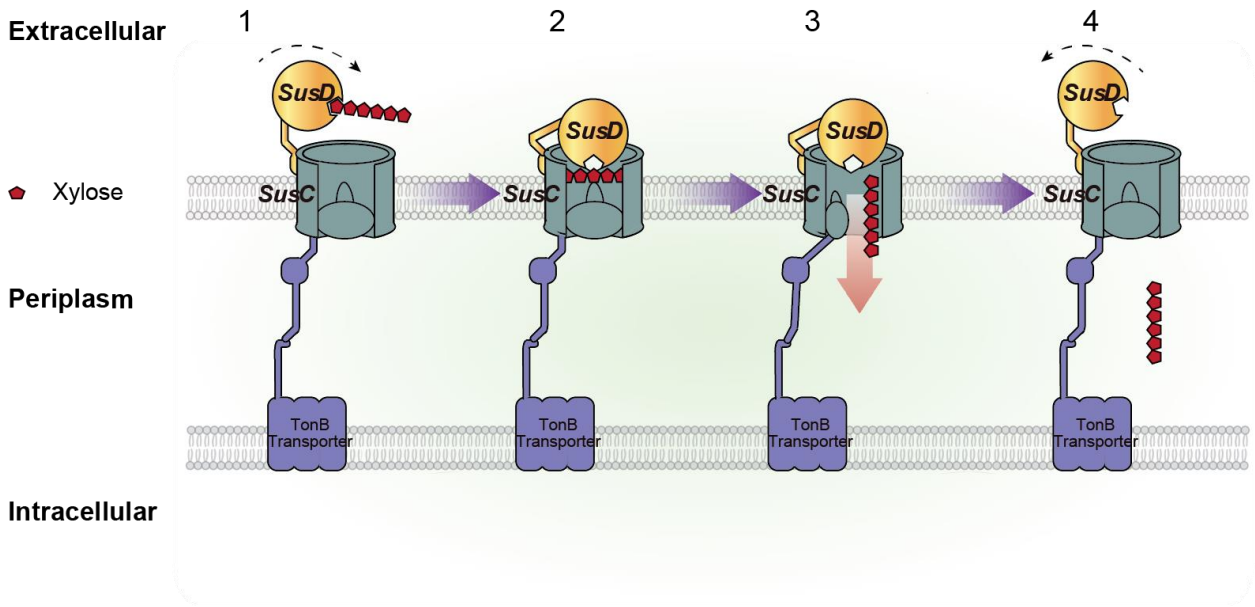


Figure 1.4 - **SusC/SusD pair works with a hinge like mechanism.** 1- SusD binds the oligosaccharides on the extracellular region and in a hinge like mechanism moves towards SusC channel. 2- SusD inserts the oligosaccharides within SusC channel, interacting with SusC plug. 3- Interaction of the TonB dependent transporter with SusC and substrate interaction causes a conformational change in SusC allowing the plug to move out of the channel. 4- Oligosaccharide enters the periplasmic region and SusD returns to its original position.

known to stabilize the substrate to allow the degradation by GH situated peripherally at the outer membrane.

Bacteroides thetaiotaomicron. This predominantly commensal bacteria, *Bacteroides thetaiotaomicron*, which was initially isolated from fecal samples of 20 healthy male Japanese-Hawaiians, resides in the human lower GIT (Moore 1974). This species is capable of utilizing a wide range of monosaccharides and polysaccharides such as chondroitin sulfate, laminarin and starch, and is an obligate anaerobe (Kotarski and Salyers 1984). Taking advantage of a wide array of polysaccharide utilization loci, this organism is also able to degrade a wide range of polysaccharides that evade host digestion as well as several host-derived glycans (Hooper, Midtvedt et al. 2002). Moreover, *B. thetaiotaomicron* possesses one of the largest representation of glycoside hydrolases from all sequenced bacteria, including amylases, β-glucosidases, α-mannosidases and 20 other activities (Xu, Bjursell et al. 2003). These glycoside hydrolases are mainly encoded in polysaccharide utilization loci, designated for the utilization of specific polysaccharides, in which are also encoded a hybrid two-component system (HTCS), carbohydrate-binding proteins and membrane transporters (Xu, Bjursell et al. 2003). These gene clusters that target polysaccharides were named after the starch utilization system, the first characterized PUL, which is required for *B. thetaiotaomicron* to grow on amylose, amylopectin, pullulan and maltooligosaccharides (Anderson and Salyers 1989). The Sus system contains eight genes that govern the degradation and transport of starch oligosaccharides inside the cell. These genes are a transcriptional activator (SusR) (D'Elia and Salyers 1996), with extracellular components that bind and degrade polysaccharides (SusDEFG) for transport into the periplasm by a TonB-

dependent β-barrel porin (SusC) (D'Elia and Salyers 1996, Shipman, Berleman et al. 2000, Cho and Salyers 2001), neopullulanase (SusA), α-glucosidase (SusB), and α-amylase activity (SusG) (Reeves, DElia et al. 1996, Shipman, Berleman et al. 2000). Recently, a novel PUL found in *B. thetaiotaomicron* demonstrated its ability to degrade rhamnogalacturonan-II, by degrading 20 different glycosidic linkages, into fermentable sugars. Highly adaptable to its environment, *B. thetaiotaomicron* may take advantage of its 101 potential PULs systems to adapt and stably colonize the human lower GIT (Xu, Mahowald et al. 2007).

Bacteroides fragilis. *Bacteroides fragilis* is the most common isolate in clinical infections by *Bacteroides*, although it only comprises about 0.5% of the human GIT microbiome, and thereby considered the most virulent *Bacteroides* species (Wexler 2007). Although, there is a large number of paralogous genes between *B. fragilis* and *thetaiotaomicron*, comparative genome analysis suggests a different microenvironment based on specific adaptations of each genome (Wexler 2007). *Bacteroides fragilis* has the ability to alter the surface of the mucosal layer using low concentrations of O_2, giving this organism a niche more related to the surface of the mucosa. This ability to use low concentrations of O_2, through oxidation of NADH by cytochrome *bd* oxidase, may play a key role during infection by allowing *Bacteroides fragilis* to grow in oxygenated tissue before forming an abscess (Baughn and Malamy 2004). Despite *Bacteroides fragilis* being an opportunistic pathogen, its core relationship with the mucosal layer and the production of polysaccharides that are capable of activating the T cell dependent immune system, plays a fundamental role in the development of the host immune system. Moreover, these polysaccharides, especially polysaccharide A, has anti-inflammatory

properties and can be delivered to the host by outer membrane vesicles. (Mazmanian and Kasper 2006). This tight association of *Bacteroides fragilis* to the mucosal tissue, demonstrated by an elegant PUL named commensal colonization factor, allows *B. fragilis* to colonize the crypts, preventing subsequent colonization by other microbes (Lee, Donaldson et al. 2013).

***Bacteroides ovatus*.** This prevalent member of human GIT, *Bacteroides ovatus*, possess a wide variety of PULs related to the degradation of dietary fiber such as arabinoxylan, xylan, arabinan, xyloglucan and others. Compared with *Bacteroides thetaiotaomicron*, this organism is more specialized to utilize plant cell wall, such as hemicellulose and pectic substrates (Martens, Lowe et al. 2011). Recently, the complete genome sequence of this colonic bacterium helped shed light on several aspects of its metabolic potential, specifically related to dietary fiber degradation (Rogowski, Briggs et al. 2015). *Bacteroides ovatus* is one of the most versatile *Bacteroides* in the colonic environment (Macfarlane, Hay et al. 1990), possessing over 100 predicted PULs (Terrapon, Lombard et al. 2015). For instance, this organism exhibits a total of 63 genes upregulated at least 5-fold when grown in wheat arabinoxylan as the only carbon source compared to xylose growth, demonstrating a complex variety of CAZy genes for the degradation of arabinoxylan (Zhang, Chekan et al. 2014). In addition, *Bacteroides ovatus* harbors a PUL for the degradation of xyloglucans, a rare metabolism, employed by this organism to fill a niche in the human GIT. Due to a high presence of xyloglucans and arabinoxylan in the human diet of cereal grains such as quinoa and wheat, *Bacteroides ovatus* may serve as a key member of the human colonic microbiome (Larsbrink, Rogers et al. 2014).

Bacteroides intestinalis. Isolated from fecal samples of healthy Japanese subjects, *Bacteroides intestinalis* DSM 17393 possesses one of the largest arrays of carbohydrate-related enzymes in the *Bacteroides* species (Terrapon, Lombard et al. 2015). Furthermore, this bacterium has evolved a vast region on its genome directed at xylan degradation, and therefore underscoring the importance of this polysaccharide to this organism. Previous transcriptome analysis of this bacterium grown in wheat arabinoxylan compared to xylose:arabinose, demonstrated up-regulation >5-fold of nearly 70 distinct genes organized in two separate polysaccharide utilization loci (PULs) (Zhang, Chekan et al. 2014). A recent study demonstrated the presence of five different endoxylanases present in those PULs, associated with several accessory enzymes targeting the full depolymerization of arabinoxylan into fermentable sugars (Wang, Pereira et al. 2016). The ability of this bacterium to degrade recalcitrant polysaccharides, together with other polysaccharide primary degraders such as *B. cellulosilyticus* and *B. ovatus*, provides nutrients for the host as well as cross feeding among the colonic microbial communities. Based on the recent recognition of xylan degradation in the gut as a major process, *B. intestinalis* may be seen as a major player in degradation of polysaccharides and in the microbial community.

1.3 Plant cell wall structure

Polysaccharides from plant cell walls are the most abundant biomass on Earth and it is an important source of dietary glycans consumed by humans (Chesson 1985, Sun and Cheng 2002). These dietary polysaccharides, with the exception of starch, are not metabolized by the host due to the lack of enzymes to breakdown the glycosidic bonds in the polymers. Therefore, most of the dietary fiber escapes host metabolism and

absorption in the small intestine, reaching the lower portion of the gastrointestinal tract and allowing microbial fermentation. The main components of the plant cell wall are: cellulose (40.6% to 51.2%), hemicellulose (28.5% to 37.2%), and lignin (13.6% to 28.1%) (Sun and Cheng 2002). Previous research on the leaves of perennial grasses and ryegrass shed light onto the composition of the plant cell wall. The primary structural polysaccharides are, in decreasing concentration, cellulose, glucurono-arabinoxylan, xyloglucan, rhamnogalacturonan, two mixed-linkage glucans (1-3 and 1-4), and galactans. The proportion of individual polysaccharides varies based on cell types (Chesson 1985).

Cellulose. Containing the largest storage of glucose in the biosphere, cellulose is the main structural component of plant cell walls. This polymer is composed of glucose monomers with a β-1,4 glycosidic linkage and presents a high degree of crystallinity. Due to this crystallinity and the stability of β-1,4 linkages, cellulose is highly recalcitrant to enzymatic degradation (Richard Wolfenden 1998). Furthermore, the linkages between the glucose monomers allow an extensive hydrogen bonding network. In the case of amorphous cellulose, this hydrogen bond network seems to be disrupted during cellulose synthesis, diminishing stacking of the cellulose polymers and thus resulting in greater susceptibility to enzymatic attack (Pavlostathis, Miller et al. 1988). Cellulose contains other cross-linked hemicelluloses such as β-glucans and xyloglucans, intertwining the cellulose microfibrils. Although crystalline cellulose requires few glycoside hydrolase families for their depolymerization into monomers, most bacteria cannot utilize crystalline cellulose (Liming Xia 1999, Thygesen, Thomsen et al. 2003).

Hemicellulose. Hemicelluloses are a heterogeneous group of polysaccharides and is the second most abundant component of renewable biomass, such as sugarcane bagasse and wheat straw (Sun, Tian et al. 2012). Hemicellulose is composed of several different polysaccharides such as xylan, xyloglucans, mannans, and glucomannans. The most prevalent hemicellulose polysaccharide, xylan, can be loosely categorized into homoxylans, glucuronoxylans, arabinoxylans, and arabinoglucuronoxylans. The common feature of xylan polysaccharides is the xylose backbone linked by β-1,4 glycosidic linkages. The main xylan that composes the hemicellulose of cereals such as rye and wheat, arabinoxylan, is structured by a xylose backbone substituted with arabinose residues. Moreover, arabinoxylans may contain ferulic acid attached to O-5 of some arabinofuranosyl residues, as well as acetyl group at O-3 of xylose residues. These ferulated residues can also be cross-linked to lignin, rendering a recalcitrant characteristic to enzymatic degradation. The diversity of xylan structure, based on different side chains and substitutions, requires an equally diverse enzymatic array to fully degrade xylan into fermentable sugars. Depending on the diverse structure of xylan a mixture of the following different enzymatic activities are required: endo-1,4-β-xylanases (EC 3.2.1.8), β-D-xylosidases (EC 3.2.1.37), α-L-arabinofuranosidases (AFs) (EC 3.2.1.55), α-glucuronidases (EC 3.2.1.139), acetyl xylan esterases (EC 3.1.1.72), and ferulic/coumaric acid esterases (EC 3.1.1.73) (Dodd and Cann 2009). Therefore, understanding complex xylan degradation requires a deep analysis of the enzymes needed for its hydrolysis, ultimately providing an understanding of dietary fiber degradation and human gastrointestinal health.

Lignin. Lignin, a major structural non-polysaccharide polymer, is composed of three different phenylpropane monomer units named *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. This complex branched 3-D phenolic polymer structure confers rigidity and hydrophobicity to plant cell walls. Moreover, the random polymerization of these three monomers generates a highly recalcitrant structure, inhibiting enzymes from accessing cellulose and hemicellulose. The composition, structure, and amounts varies throughout different species, tissues, and even in the same organism (Pandey and Kim 2011). Lignin is deposited in differentiated cells that function as support for tissues responsible for water conduction (Bonawitz and Chapple 2010). The complexity of the structures of lignin is one of the most limiting factors of hemicellulose and cellulose degradation by microorganisms.

1.4 Enzymatic deconstruction of plant cell wall

Due to the vast structural diversity of plant cell walls, it requires an equally diverse set of enzymes for the degradation of these polysaccharides. Glycoside hydrolases cleave glycosidic linkages found in the polysaccharide structures, and carbohydrate esterases cleave ester-bound moieties. Much is known about the enzymatic functionalities required to degrade polysaccharides, and numerous enzymes have been identified from a variety of microbial sources (Cantarel, Coutinho et al. 2009).

The endoxylanases. The xylan backbone is composed of xylose residues linked by β-1,4 glycosidic linkages. Thus, one of the most important enzymatic activities towards deconstruction of the xylan backbone is the endo-1,4-β-xylanase activity. These enzymes are responsible for cleaving the β-1,4 glycosidic linkage between xylose residues. Several different GH families are known to have endoxylanase activity such as GH families 5, 7,

8, 10, 11, and 43, based on the amino acid sequence, structural fold and catalytic mechanism (Cantarel, Coutinho et al. 2009). The end-products from the reaction of endoxylanases with xylan substrates vary among the different families (Wang, Pereira et al. 2016). Xylanases possess binding sites for xylose residues termed subsites with bond cleavage occurring between sugar residues at the -1 subsite and the +1 subsite, non-reducing end and reducing end, respectively, of the polysaccharide substrate (Davies GJ 1997). Although both GH10 and GH11 endoxylanases are able to bind to ramified xylo-oligosaccharides (Fujimoto, Kaneko et al. 2004, Pell, Taylor et al. 2004, Vardakou, Flint et al. 2005, Vardakou, Dumon et al. 2008), GH10 enzymes can accommodate +1 subsite modifications on the xylose (Kaneko, Ichinose et al. 2004, Pell, Szabo et al. 2004), while GH11 enzymes can accommodate linkages on xylose at the +2 subsite (Vardakou, Dumon et al. 2008).

The α-L-arabinofuranosidases. Arabinose is a common substitution in arabinoxylan side chains. Members of the GH 43, 51, 54, and 62 families (Henrissat and Bairoch 1996), the α-L-arabinofuranosidases (AFs), have the function of cleaving arabinose side chains of the xylose backbone. This enzymatic activity acts synergistically with endoxylanases for efficient degradation of arabinoxylan. The α-L-arabinofuranosidases are ubiquitous and have been detected members in the different domains or life, including members of the filamentous fungi such as *Aspergillus nidulans* (Ramon, vd Veen et al. 1993), bacteria, such as *Bacteroides ovatus* and *Bacteroides intestinalis* (Whitehead 1995, Zhang, Chekan et al. 2014), and archaea, such as *Pyrodictium abyssi* (Andrade, Aguiar et al. 2001). The α-L-arabinofuranosidases are grouped into three different categories: type A enzymes are active on short arabino-

oligosaccharides; type B enzymes possess a wider range of substrates such as arabinoxylan, arabinan, and arabino-oligosaccharide; Type C enzymes, arabinofuranohydrolases, are mainly active on arabinoxylan substrates (Saha 2000).

β-xylosidases. The 1,4-β-D-xylosidases are enzymes responsible for catalyzing the hydrolysis of the xylose residues from the non-reducing ends of xylo-oligosaccharides. Furthermore, based on the CAZy database (Cantarel, Coutinho et al. 2009), these enzymes are grouped into 11 different GH families (GH1, GH3, GH30, GH31, GH39, GH43, GH51, GH52, GH54, GH116, and GH120). Largely used in industrial processes, these enzymes are key for increasing the quality and improving the baking of bread dough (Dornez, Gebruers et al. 2007). They have also been used in the production of D-xylose for xylitol manufacturing (Polizeli, Rizzatti et al. 2005). Most glycoside hydrolases, including 1,4-β-D-xylosidases, employ one of the two accepted reaction mechanisms, i.e., the inversion and retaining mechanisms (Koshland 1953). In the inverting mechanism, the β-anomeric configuration of the substrate non-reducing end is released with an α-xylose configuration, while through the retaining mechanism, the reaction occurs in two transition states. During the first transitional state, an ester bond is formed between the nucleophile amino acid (asparagine or glutamic acid) and the target C1 carbon in the xylose residue. In the second transitional state, a dexyloxylation happens, hydrolyzing the ester bond and releasing β-xylose (Jordan and Wagschal 2010). The mechanisms are illustrated in fig. 1.2 above.

Carbohydrate esterases. The carbohydrate esterases (CEs) are a class of enzymes that catalyzes the removal of O- or de-N-acylation of esterified decorations from carbohydrates (Cantarel, Coutinho et al. 2009). Currently, there are 15 different families

of CEs in the CAZy database, while having a wide range of activities. Therefore, the ability of these enzymes to remove side chain decorations from the polysaccharide backbone works synergistically with the ability of glycoside hydrolases to depolymerize complex glycans. One of the most diversified CE families, CE1, is composed of more than 4000 proteins. This family has a wide range of activities such as acetylxylan esterase (EC 3.1.1.72), feruloyl esterase (EC 3.1.1.73), and *O*-acyltransferases (EC 2.3.1.20). The proteins that have their three dimensional structures solved, assume an α/β hydrolase conformation with a central β-sheet with eight or nine β-strands surrounded by α-helices. The activities of these enzymes are key to efficient degradation of decorated complex glycans, highlighting their relevance especially for arabinoxylan degradation.

1.5 Effects of metabolites on host health and immune system

The human body is inhabited by a vast number of microbes, which based on recent estimates could reach around 3.8×10^{13} in adult males and 4.4×10^{13} in adult females. Moreover, most of these microbes are in the distal portion of the gastrointestinal tract. Analysis of the large microbial community shows that the human gut is colonized mainly by two phyla, the Firmicutes and the Bacteroidetes, with lower numbers of the Proteobacteria phylum (Hold, Pryde et al. 2002, Eckburg, Bik et al. 2005). Within this large microbial community is a wide metabolic potential. Current metagenomics sequencing of fecal samples has identified more than 3×10^6 unique microbial genes, outnumbering the metabolic potential of the host by about 150-fold. Furthermore, it is estimated that each individual is colonized by at least 160 different species (Qin, Li et al. 2012). While most of these bacteria are still regularly described as commensal bacteria, i.e. not harmful nor beneficial, this host-microbe interaction seems to be highly mutualistic.

This relationship benefits both the microbes, such as access to nutrients, and the host, with degradation of complex nutrients and synthesis of essential metabolites such as vitamins (Cummings and Macfarlane 1997). The current estimate is that about 10% of metabolites found in the blood stream are derived from the gut microbiome (Wikoff, Anfora et al. 2009). Therefore, the host and microbes have a finely tuned interaction, which perturbed can affect the metabolic and immune systems, leading to metabolic syndromes or chronic inflammation such as obesity and atherosclerosis (Cox and Blaser 2013, Jonsson and Bäckhed 2017).

Host microbiome and obesity. Based on data from the National Health and Nutrition Examination Survey (NHANES), more than 1 out of 3 adults are considered obese in the United States. Obesity is one of the main risk factor for metabolic syndromes, which can lead patients to type 2 diabetes and cardiovascular problems (Flegal, Kruszon-Moran et al. 2016). Recent studies demonstrated that the gut microbiome is an important factor that leads to obesity, in combination with diet and the host genetic background (Cox and Blaser 2013). Although the microbiome composition greatly vary between hosts, several studies report a relative increase in the Firmicutes:Bacteroidetes ratio in overweight and obese animals and humans (Ley, Backhed et al. 2005, Ley, Turnbaugh et al. 2006). It is hypothesized that high fat-low fiber diets, commonly described as a westernized diet, are highly rich in calories from sucrose and unsaturated fat, selecting a less diversified microbiota that utilize these nutrients more efficiently, such as the Firmicutes. In contrast, a fiber-rich diet promotes Bacteroidetes, which are efficient primary degraders of dietary fibers (De Filippo, Cavalieri et al. 2010, Wu, Chen et al. 2011). Despite the large variability among individuals, the metabolic capacity of healthy

microbiomes are strikingly similar, while obese individuals also have a great similarity among their microbiome's biochemical capacities even with a less diverse community (Kurokawa, Itoh et al. 2007, Qin, Li et al. 2012).

Furthermore, microbiome transplantation from obese mice to its germ free counterparts generates a significant weight gain compared to a similar transplantation from lean mice. This clearly demonstrates that the microbiome plays a key role in controlling body weight and obesity (Turnbaugh, Backhed et al. 2008). In another small study, the transplantation of a microbiome from a lean subject to a type 2 diabetes obese patients did not have any significant weight reduction, although this led to an increase in insulin sensitivity 6 weeks after the transplant (Vrieze, Van Nood et al. 2012). These results, although promising, still require longitudinal studies in a larger scale population to identify the mechanism by which fecal matter transplantation (FMT) regulates metabolic syndrome disorders.

Role of microbiota in type 2 diabetes. One of the main causes of type 2 diabetes is a systemic chronic inflammation of insulin-producing β-cells of the pancreatic tissue and white adiposity tissue itself. This inflammation leads to a decreased sensitivity to insulin in white adipose cells as well as decrease insulin production by the pancreas β-cells (Donath, Dalmas et al. 2013). Due to the chronic inflammation component of type 2 diabetes, the current hypothesis of the role of the gut microbiome is a diminished metabolic capacity to ferment dietary fiber and production of short-chain fatty acids (Fukuda, Toh et al. 2011, Wrzosek, Miquel et al. 2013). Short chain-fatty acids increase the integrity and productions of the intestinal mucosa, thus, lower production of these can cause an increased permeability of the intestinal tight junctions allowing bacteria from the

gut to permeate into the blood stream (Willemsen, Koetsier et al. 2003, Gaudier, Jarry et al. 2004, Fukuda, Toh et al. 2011, Wrzosek, Miquel et al. 2013, Macia, Tan et al. 2015). Immune and adipose cells recognize the pathogen-associated molecular patterns, such as lipopolysaccharide, through pattern recognition receptors increasing the production of pro-inflammatory cytokines. Moreover, LPS can be transported through chylomicrons and lipoprotein complexes from the lumen to the plasma and blood. These pro-inflammatory cytokines generate a low, chronic, and generalized inflammation level that potentially can exacerbate insulin resistance and also its low production (Khan, Nieuwdorp et al. 2014).

Atherosclerosis and human microbiome. One of the main risks of cardiovascular disease, atherosclerosis, is a disproportionate accumulation of cholesterol phospholipids and triglycerides associated with low-density lipoproteins (LDL) in the arteries. The oxidation and enzymatic modification of this has pro-inflammatory properties that recruit macrophages that uptake lipids deposited. This leads to a feedback loop of pro-inflammatory cytokines productions, attracting more macrophages to the site and causing inflammation (Moore, Sheedy et al. 2013). Although atherosclerosis is caused by a chronic inflammation of arteries, the mechanisms of this inflammation are different from type 2 diabetes. The gut microbiome influences this effect by the degradation of phosphatidylcholine (PC) from the diet. PCs, a class of phospholipids that has a choline as a head group, are one of the major component of biological membranes and are abundantly found in eggs, cheese, and red meat (Jonsson and Backhed 2017). Processing of PC by the gut microbiome generates trimethylamine, readily absorbed by the host and converted into trimethylamine N-oxide (TMAO). High blood levels of TMAO have been associated with hearth attack, stroke, and atherosclerosis (Wang, Klipfell et

al. 2011, Zhu, Gregory et al. 2016). Moreover, L-carnitine, an amino acid found in red meat is processed to TMAO by gut bacteria, leading to atherosclerosis in mice, and is associated with cardiovascular risks (Koeth, Wang et al. 2013).

Microbiome metabolites and host immune system. The gut microbiome was believed to evade the immune system to be able to thrive in the gastrointestinal track, however recent studies has demonstrated that the host-microbe interactions goes far beyond this aspect. Instead, this complex system has a large interaction of the microbiome, immune cells, and intestinal epithelium to actively reach the gastrointestinal tract homeostasis (Matsumoto, Kibe et al. 2012). The key role in this intrinsic communication is the bacterially derived metabolites, which act as signals for the immune system and intestinal functions (Arpaia, Campbell et al. 2013, Ferreira, Vieira et al. 2014, Kim, Qie et al. 2016).

Natural phenolic compounds. Recently there has been an increased interest in using the natural antioxidant properties of plant phenolic compounds for their therapeutic applications. These plant antioxidant molecules can act as reducing agents, metal chelators, and free radical scavengers (Ohnishi, Matuo et al. 2004). There is a high concentration of these phenolic compounds present in fruits, vegetables, and grains that provide efficient antioxidant effects (Figure 1.4) (Gani, Sm et al. 2012). Out of these phenolic compounds, ferulic acid is the most abundant in plants, found in high concentrations in grains used in human diet such as wheat bran and corn bran (Gani, Sm et al. 2012). Currently, there are several studies demonstrating the pharmacological applications of ferulic acid in the attenuation of human diseases such as diabetes, high cholesterol, and neurological disorders

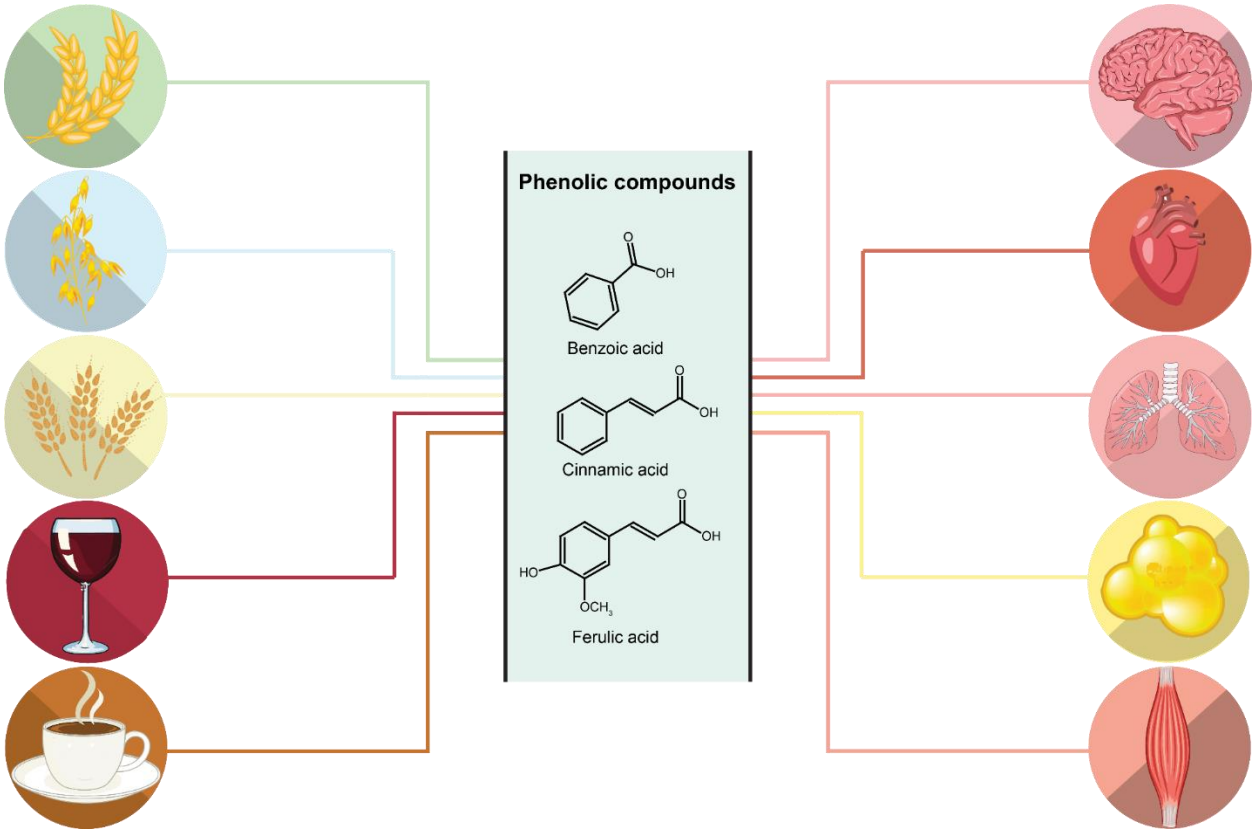


Figure 1.5 - **Phenolic compound sources and benefits to the host.** Representation of dietary ferulic acid sources from whole cereal grains (wheat, oat, and rye), alcoholic beverages (red wine, beer), and coffee. Benefits to the host are represented by reduced inflammation of brain, heart, lungs, decrease insulin resistance and lower glycemic index in adipose and muscular tissues.

(Ohnishi, Matuo et al. 2004, Jin Son, C et al. 2010, Dong, Kuan et al. 2015). These potential therapeutic proprieties may be attributed, in part, to its high capability to sequester free radical molecules.

Short chain fatty acids as signals for the immune system. Extensive research has been conducted on the effects of short-chain fatty acids on the host. These metabolites are fatty acids with a backbone compose between 1 to 6 carbon atoms, which are largely produced by bacterial fermentation of plant-related polysaccharides that evade the host digestion. The main SCFA produced by microbes in the distal gut are acetate, propionate, and butyrate, which represent about 3-10% of the energy intake of the host. Despite this large influence of SCFA to the host nutrition, one of the main aspects is the signaling to the immune system and intestinal epithelium integrity (Kalina, Koyama et al. 2002, Macia, Tan et al. 2015). SCFA act as anti-inflammatory molecules in the gastrointestinal tract, specially, with a protective role to inflammatory bowel disease (IBD), as largely demonstrated in mice model studies (Maslowski, Vieira et al. 2009, Furusawa, Obata et al. 2013, Smith, Howitt et al. 2013). Similar to these studies, it has been demonstrated that IBD is correlated to a decrease in SCFA producing bacteria and that dietary enrichment of these bacteria has a role in ameliorating IBD symptoms (Kanauchi, Suga et al. 2002, Sokol, Seksik et al. 2009). The more efficient plant degraders, i.e., the Bacteroidetes are responsible for most of the acetate and propionate produced, while the Firmicutes are the main producers of butyrate (El Kaoutari, Armougom et al. 2013, Levy, Thaiss et al. 2016). The short chain fatty acids also play a large role in inter-species crossfeeding. The Firmicutes, for example *Faecalibacterium prausnitzii*, can utilize

acetate produced by the Bacteroidetes as a substrate to produce butyrate (Duncan, Holtrop et al. 2004).

The pool of SCFA produced by the gut microbiome are available for metabolism by the host, although each have different routes in the host metabolism. While butyrate, major source of energy for the colonic enterocytes, is readily consumed by the mucosal lining, acetate and propionate have a different fate (Donohoe, Collins et al. 2012). Acetate and propionate cross the intestinal epithelium with little metabolism by the host and reach the liver, where propionate is degraded and acetate is released into systemic circulation (Bloemen, Venema et al. 2009, Morrison and Preston 2016).

Nearly every aspect of the gastrointestinal immune system is influenced by the anti-inflammatory properties of SCFA. Several studies in a colitis mouse model demonstrated that SCFA is able to diminish the effects of the disease by the induction of the repair cytokine interleukin (IL-) 18 in the epithelial cells of the GIT (Kalina, Koyama et al. 2002, Macia, Tan et al. 2015). Moreover, SCFA increase the production of mucin by goblet cells as well as modify the tight junctions (Willemsen, Koetsier et al. 2003, Gaudier, Jarry et al. 2004, Fukuda, Toh et al. 2011, Wrzosek, Miquel et al. 2013). Therefore, SCFA help promote the integrity of the intestinal lining.

Short-chain fatty acids have also been demonstrated to affect the cells of the innate immune system. The exposure to SCFA in the intestinal mucosa activates FFAR2, essential for the intestinal mucosa neutrophil chemotaxis, reducing the production of pro-inflammatory cytokines of LPS-stimulated neutrophils (Le Poul, Loison et al. 2003, Sina, Gavrilova et al. 2009, Vinolo, Rodrigues et al. 2011).

Regulatory T cells (Treg cells) have an important role in diminishing the inflammatory responses in the gastrointestinal tract. Therefore, one of the most relevant anti-inflammatory aspects of SCFA is the promotion of Treg cells, which modulates the activity of effector T cells. Acetate and propionate have been demonstrated to stimulate existing Tregs in the colon by activation of FFAR2, while butyrate and propionate induce the differentiation of naïve T cells into Treg cells by inhibition of HDAC in dendritic cells and T cells (Arpaia, Campbell et al. 2013, Furusawa, Obata et al. 2013, Smith, Howitt et al. 2013).

Amino acids metabolites have immunomodulatory effects. The arginine derivatives, putrescine, diamine, spermidine, and spermine are high reactive polyamines produced and secreted by bacteria in the GIT (Michael 2016). Polyamines are present in every cell and have an important role in gene expression and cell proliferation, with its endogenous production tightly regulated by the host (Miller-Fleming, Olin-Sandoval et al. 2015). Several studies have shown that supplemented polyamines have a role in intestinal mucosa integrity as well as the immune system (Löser, Eisel et al. 1999, Perez-Cano, Gonzalez-Castro et al. 2010). Furthermore, spermine directly influence monocytes and macrophages by inhibiting the pro-inflammatory cytokines produced in response to LPS stimuli (Zhang, Borovikova et al. 1999, Zhang, Luo et al. 2012).

The three aromatic amino acids, tryptophan, tyrosine, and phenylalanine can be metabolized into indole substrates as well as propionic acid derivatives, respectively. Recent work has demonstrated the pathways by which *Clostridium sporogenes* converts different aromatic amino acids into metabolites found in the host serum (Dodd, Spitzer et al. 2017). Indolepropionic acid (IPA) is a potent radical-scavenging molecule and has

been shown to increase the stability of the intestinal barrier. Several other indole derivatives also act as ligand to the aryl hydrocarbon receptor in the host cells (Wikoff, Anfora et al. 2009) and increase resistance to infection by *Listeria monocytogenes* and *Citrobacter rodentium* in animal models.

Polysaccharide A acts as anti-inflammatory molecule. The polysaccharide A (PSA) produced by *Bacteroides fragilis* is able to induce secretion of the anti-inflammatory cytokine IL-10 by CD4 cells and promotes differentiation of T cells into Th1 cells (Cobb, Wang et al. 2004, Dasgupta, Erturk-Hasdemir et al. 2014, Jonsson and Backhed 2017). This production of IL-10 is commonly mediated primarily by dendritic cells, which internalize the PSA and present to naïve T cells. Another mechanism of action to this anti-inflammatory response is the activation of TLR 2 express in dendritic cells that lead to expression of cytokines that induce production of IL-10 by T cells (Mazmanian, Round et al. 2008, Dasgupta, Erturk-Hasdemir et al. 2014).

CHAPTER 2: FUNCTIONAL ANALYSIS OF AN ESTERASE-ENRICHED POLYSACCHARIDE UTILIZATION LOCI

2.1 Introduction

The human gastrointestinal microbiome, through co-evolution and host selection, has generated a complex and stable community capable of modulating host physiology (Lozupone, Stombaugh et al. 2012). Changes in the diversity of microbiome populations are associated with disease states for the host (Karlsson, Tremaroli et al. 2013, Louis, Hold et al. 2014). *Bacteroides* are a predominant group of commensal microorganisms within the human gastrointestinal tract (GIT) responsible for a network of metabolic interactions with the host and other colonic microbes. Within this communication between host and microbes, fermentation of dietary fiber plays a key role in the lower gastrointestinal tract. Arabinoxylan, a hemicellulosic polysaccharide, is one of the most prevalent structural polymers found in cereal grains, such as wheat, oats, and rye. Highly heterogeneous, this polysaccharide is mainly composed of xylose, arabinose, decorated moieties, and phenolic compounds. Recent pharmaceutical interest in natural phenolic compounds, like ferulic acid, results from potential therapeutic uses against diseases. Ferulic acid esters covalently link microbially fermentable hemicellulose to lignin, hindering its deconstruction (Jordan and Wagschal 2010). The genus *Bacteroides* constitutes primary degraders of diet-derived carbohydrates and possess an extensive array of carbohydrate degrading enzymes. Carbohydrate esterases are widely conserved within members of the human GIT, highlighting the importance of ester-bound moieties in degradation of dietary fiber (Bhattacharya, Ghosh et al. 2015). Previous studies demonstrated a large variability in the healthy gut microbial community, thus leading to

the hypothesis that a healthy microbiome is reliant on the functions that different microbes provide in its ecological niche. Moreover, metagenomics and metatranscriptomics studies provide insight into the GIT environment. However, mechanistic information for the genes is key for downstream applications. Towards this aim, understanding the activity of carbohydrate esterases will increase the knowledge of carbohydrate degradation and phenolic compounds released by members of the human colon.

In this study, we identified a polysaccharide utilization loci (PUL), conserved in several commensal bacteria from the human gut, encoding multiple carbohydrate esterases. Interestingly, this cluster seems to be regulated through the sensing of ferulated arabinoxylan substrates. Within this cluster, we analyzed the activity of several carbohydrate esterases encoded in the esterase-enriched PUL in the human colonic bacterial community member *B. intestinalis* DSM 17393. These carbohydrate esterases demonstrate variable function and specificities towards different esterified substrates. Two of the putative esterases, i.e., BACINT_01033 and the hypothetical BACINT_01040, act as feruloyl esterases, while BACINT_01038 and BACINT_01039 demonstrate acetylxylan esterase activities. Furthermore, these carbohydrate esterases demonstrate synergistic activity with several glycoside hydrolases, including enzymes from a different organism such as *Prevotella bryantii* and enzymes produced by *B. intestinalis*. This synergistic activity allows an increased activity by the carbohydrate esterases towards wheat arabinoxylan, thereby releasing larger amounts of ferulic acid and reducing sugars. Surprisingly, the hypothetical protein encoded in this PUL, BACINT_01040, exhibits promiscuous activity towards all ferulated substrates tested, compared to the very specific BACINT_01033. The broad capability of different *Bacteroides* species to remove ferulic

acid decorations from arabinoxylan rich substrates, such as wheat bran, elicited an immune response on intestinal cells and a mouse animal model. Moreover, the analysis of this PUL provided insight into carbohydrate metabolism as well as metabolic interactions between the colonic members and the human host.

2.2 Results

An esterase-enriched cluster is highly expressed during growth on complex arabinoxylan. Several different genes were highly upregulated in *Bacteroides intestinalis*, *Bacteroides cellulosilyticus*, and *Bacteroides oleiciplenus* during growth on the complex polysaccharide insoluble wheat arabinoxylan (InWAX) compared to the growth on a mixture of its component monosaccharides xylose and arabinose. Within the genes upregulated in these three *Bacteroides* species, we identified an esterase-enriched polysaccharide utilization loci (PUL) containing several esterases and glycoside hydrolases. Despite minor differences in the constitution of this PUL, the core genes related to polysaccharide degradation are present in the three different PULs upregulated in the diverse *Bacteroides* spp. For instance, in *Bacteroides intestinalis*, the genes in the PUL contained several different putative esterases encoded by BACINT_01033, BACINT_01038, and BACINT_01039, as well as putative glycoside hydrolases. The glycoside hydrolases included GH43 and GH3 families encoded by BACINT_01035, BACINT_01041, BACINT_01043, and BACINT_01042. Furthermore, this PUL also possess a pair of SusC/SusD-like proteins, responsible for the transport of oligosaccharides, and a hybrid two-component system (HTCS), predicted to regulate the transcription of PULs. Interestingly, one of the most transcribed genes in the esterase-

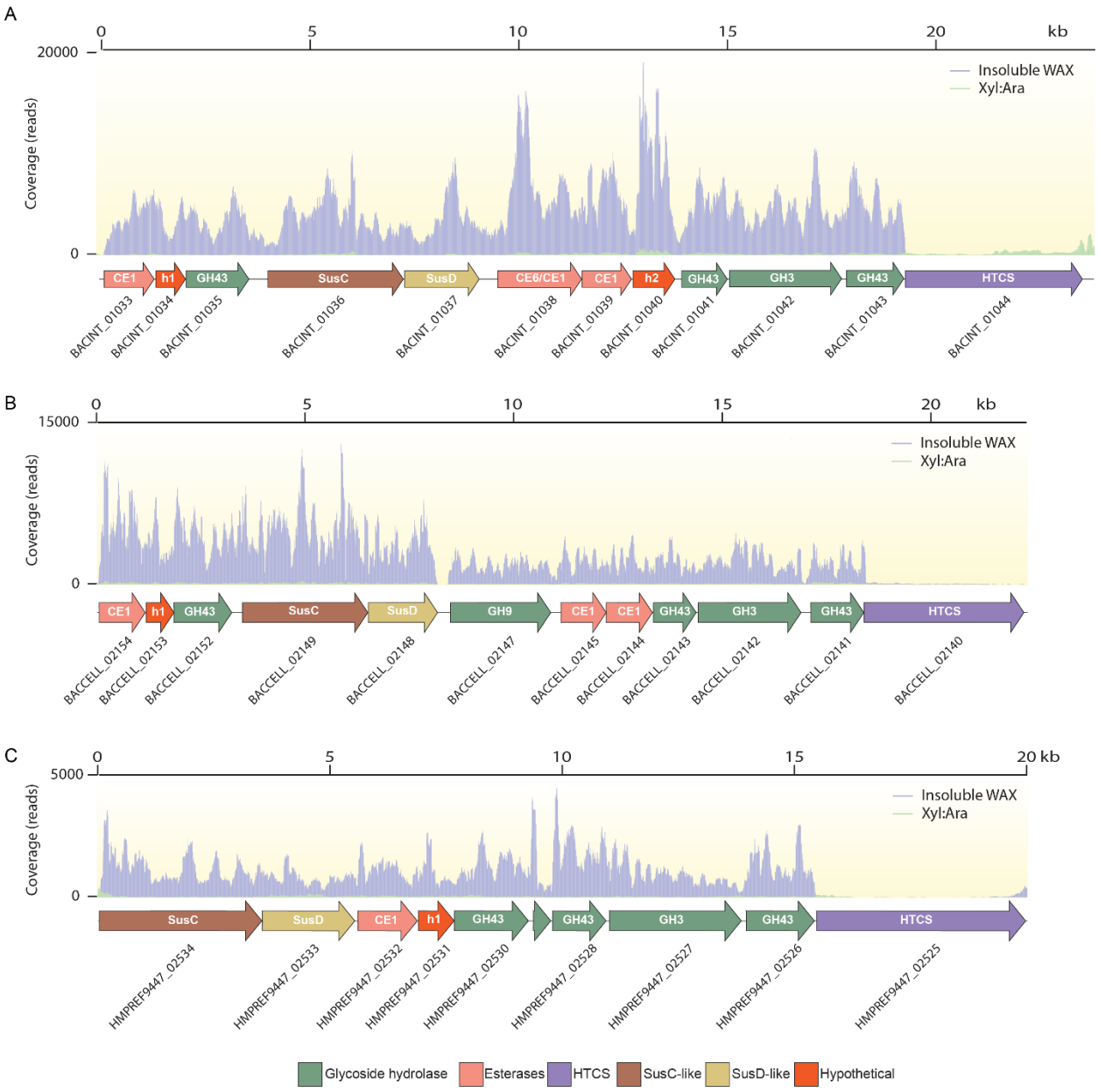


Figure 2.1 - **Comparative transcriptomics analysis of *Bacteroides* spp grown in insoluble wheat arabinoxylan (inWAX) and xylose:arabinose mixture.** Transcriptome map showing RNASeq coverage of the upregulated esterase-enriched PUL during growth on inWAX (purple) compared to monosaccharide mixture (green) in *B. intestinalis* (A), *B. cellulosilyticus* (B), and *B. oleiciplenus* (C). PUL is colored coded based on predicted functions: Hybrid two-component system (HTCS, purple), SusC/SusD-like genes (brown and yellow), glycoside hydrolases (green), esterases (salmon) and Hypothetical genes (red)

enriched cluster in *B. intestinalis* is a hypothetical protein with no function associated (Figure 2.1A). Similarly to the *B. intestinalis* transcription map, *Bacteroides cellulosilyticus* demonstrate a similar regulation pattern in its related PUL. This PUL contains three different esterases encoded by BACCELL_02154, BACCELL_02145, and BACCELL_02144. Furthermore, the *B. cellulosilyticus* PUL also contained similar glycoside hydrolases such as three GH43 (BACCELL_02152, BACCELL_02143, BACCELL_02141) and a single GH3 (Figure 2.1B).

For *Bacteroides oleiciplenus*, despite the similarities in the cluster up-regulation, the cluster in this species seems to have a lesser response to the esterified substrate in comparison to the other two species. Furthermore, the PUL in this species seems to harbor a reduced number of esterases, with HMPREF9447_02532 seeming to be the only esterase gene. Furthermore, the SusC/SusD-like pair are located at the beginning of the gene cluster, which is different from the arrangement in the two other *Bacteroides* spp. The GH43 and GH3 encoded in the other two Bacteroides spp are present in *B. oleiciplenus* and are encoded by HMPREF9447_02530, HMPREF9447_02528, HMPREF9447_02526, and HMPREF9447_02527 respectively (Figure 2.1C).

Gene homology of the esterase-enriched cluster in *B. intestinalis*. By multiple genome alignment, we were able to determine the synteny of the esterase-enriched cluster in the three *Bacteroides* species (*B. intestinalis*, *B. cellulosilyticus*, and *B. oleiciplenus*) isolated from different populations. Despite a few genes missing between the species, such as the hypothetical gene in *B. intestinalis* and the GH9 from *cellulosilyticus*, most of the core genes for polysaccharide degradation are present in the

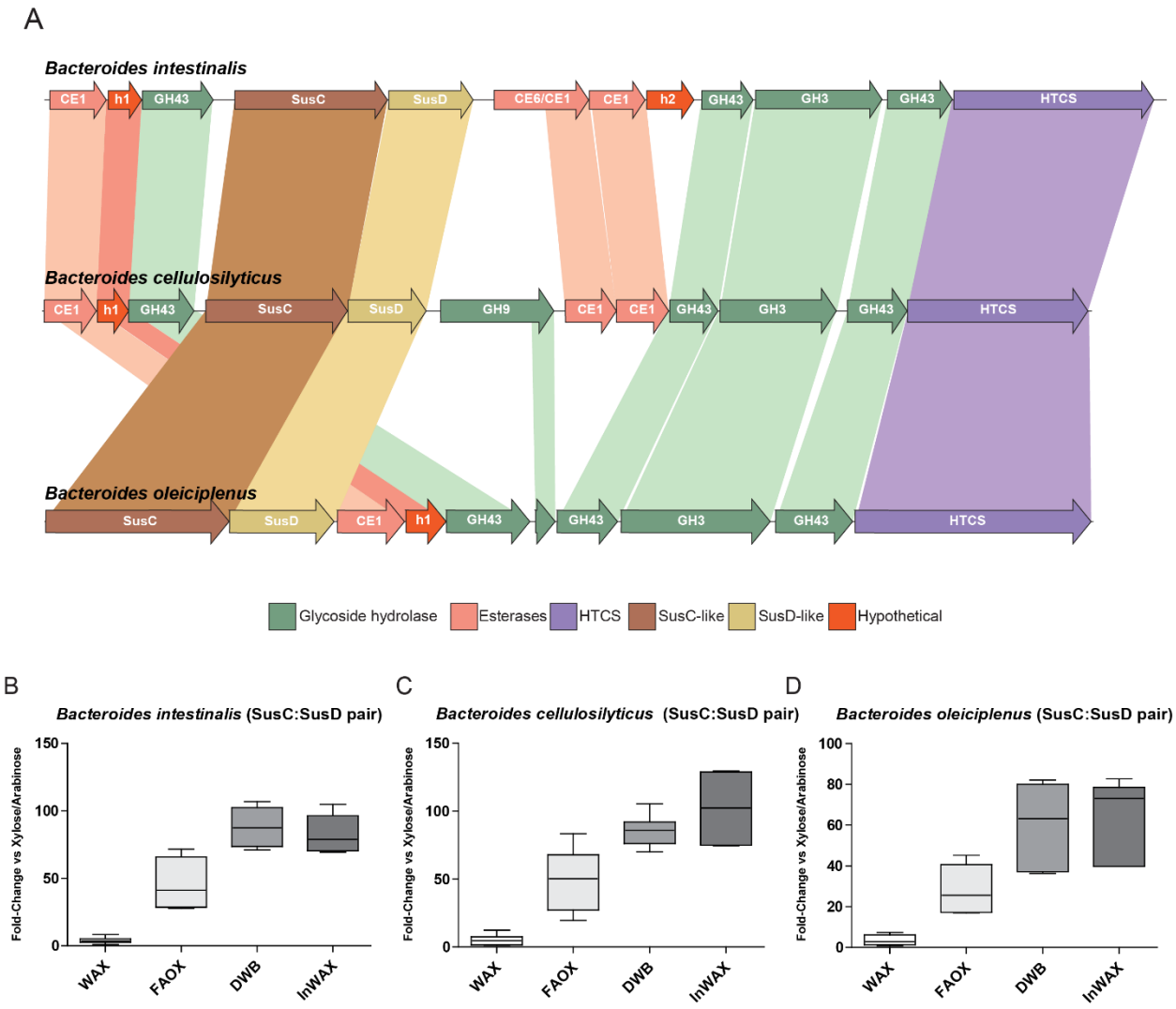


Figure 2.2 - **Esterase-enriched cluster, up regulated by ferulated arabinoxylan, is conserved throughout colonic members.** (A) MAUVE alignment of several members of the colonic environment distributed in different populations demonstrate the conservation and synteny of the esterase cluster. (B) Combined box plot of SusC/SusD RT-qPCR, normalized by 16S rRNA, in different conditions shows the differential expression of esterase-enriched cluster in B. intestinalis in different arabinoxylan substrates compared to a mixture of xylose:arabinose

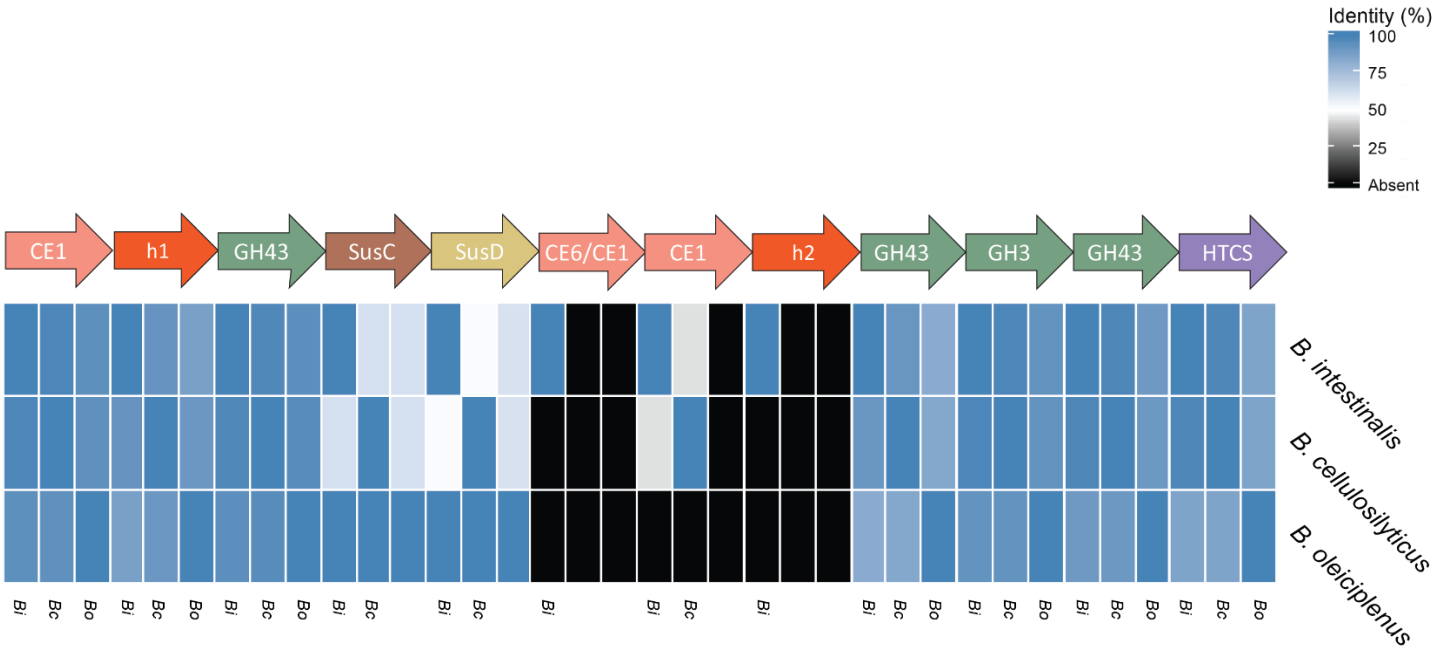


Figure 2.3 - **Esterase encoding gene cluster demonstrates high conservation in members of the human colonic *Bacteroides* spp.** Heat map depicting the cross identity of the proteins in the esterase-enriched cluster in *B. intestinalis*, *B. cellulosilyticus*, and *B. oleiciplenus*. Genes absent in the species are marked as black.

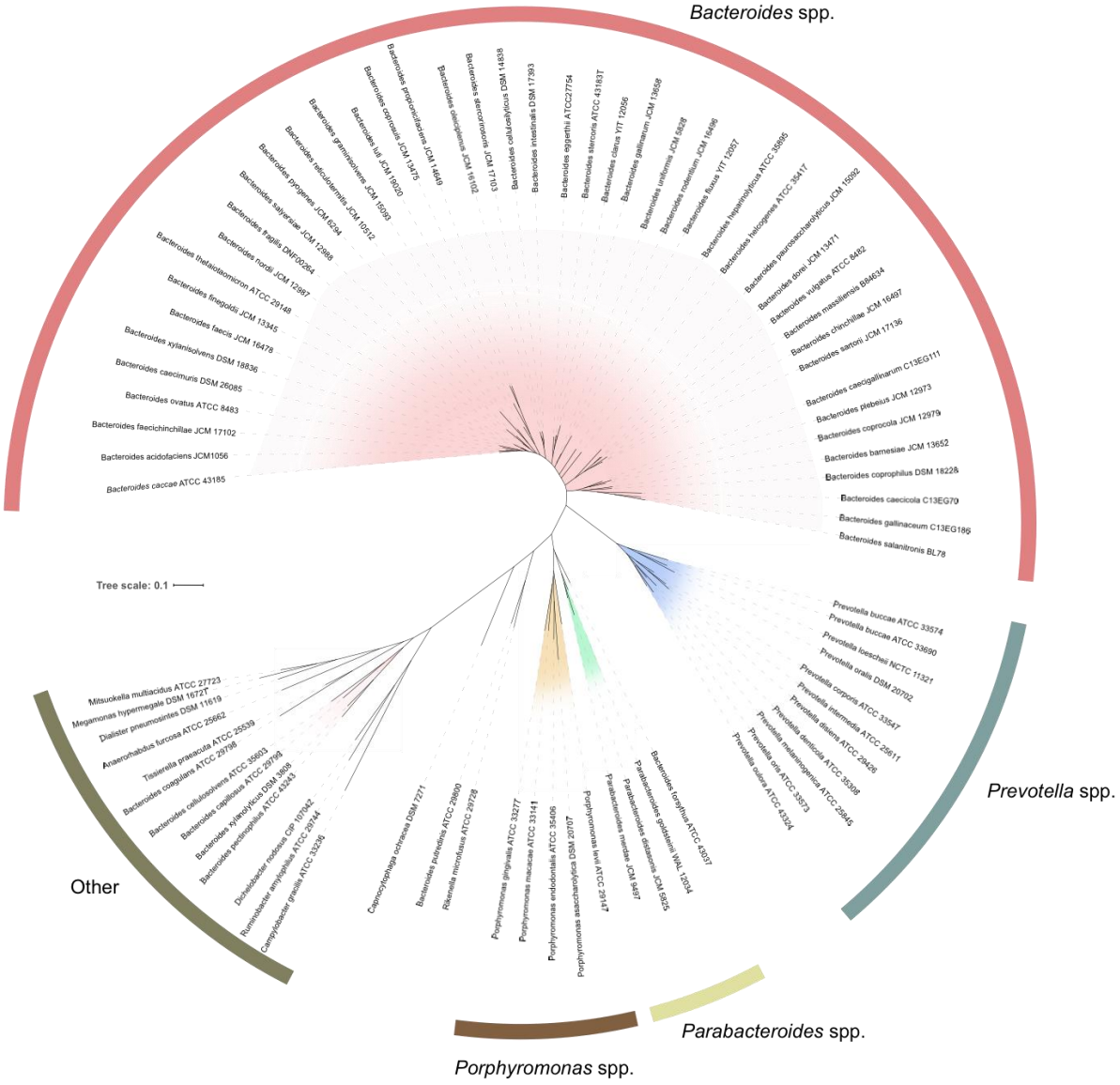


Figure 2.4 - **Phylogenetic analysis of 16S rDNA of *Bacteroides* species demonstrate close relationship of esterase-enriched cluster encoding bacteria.** Phylogenetic analysis shows clustering between *Bacteroides* (red), *Prevotella* (green), *ParaBacteroides* (yellow), and *Porphyromonas* (brown) species.

clusters. Furthermore, the clusters possess a high gene conservation when pair-wise compared. (Fig 2.2A). The synteny, combined with a similar transcription profile between the species leads to the hypothesis that this cluster is up regulated by esterified arabinoxylan. Previous studies by Zhang *et al* on transcriptional analysis of *B. intestinalis* during growth in wheat arabinoxylan did not demonstrate up-regulation of the esterase-enriched cluster. Therefore, to understand the responses of these bacteria to esterified substrates, we accessed the up-regulation of the cluster using SusC/SusD-like genes as proxy for the regulation of the cluster under different arabinoxylan-related carbohydrates compared to a mixture of xylose:arabinose. The SusC/SusD-like genes were up regulated during growth in insoluble wheat arabinoxylan (InWAX), destarched wheat bran (DWB), ferulated arabinooligosaccharides (FAOX), and only partially up regulated during growth in wheat arabinoxylan in *B. intestinalis*, *B. cellulosilyticus*, and *B. oleiciplenus* (Fig 2.2B, C, and D).

Proteins encoded by the esterase-enriched cluster are highly conserved among different *Bacteroides* species. To identify the functions of the enzymes encoded by the esterase-enriched cluster, we sought to understand the conservation of the protein sequences among three different species. Despite a variance within the SusC/SusD-like protein sequences between the different species, there is a very high conservation in the other genes products encoded by the esterase-enriched cluster, reaching up to 90% identity for GH43a and an average of ~70% across the other proteins (Figure 2.3).

***Bacteroides* species that possess the esterase-enriched cluster are closely related.** Bacteroidetes are a very diverse group, thus to understand the relationship of

the *Bacteroides* species that possess the esterase enriched cluster, a phylogenetic analysis of 16S rDNA sequence was carried out. Indeed, the presence of this cluster has evolutionary relationship between the bacteria; the phylogeny demonstrated that all three *Bacteroides* species that encoded this esterase-enriched cluster diversified from the same branch, potentially coming from a single common ancestor (Figure 2.2). This raises the hypothesis that these bacteria evolved a single cluster to degrade highly esterified substrates.

Multiple esterases and glycoside hydrolases in the esterase-enriched cluster in *Bacteroides intestinalis*. The commensal bacterium *B. intestinalis* has been demonstrated to possess significant capability in degrading arabinoxylan by up-regulating two distinct PULs (Zhang, Chekan et al. 2014). Despite the utilization of these clusters, *B. intestinalis* employ other polysaccharide utilization loci for the degradation of highly esterified arabinoxylan substrate (Fig 2.5A). Within this cluster, there are several genes encoding enzymes predicted to cleave ester linkage side chains, as well as four predicted glycoside hydrolases. The putative carbohydrate esterases are BACINT_01033 (Predicted CE1 family), BACINT_01038 (predicted CE6/CE1), BACINT_01039 (predicted CE1 family), and BACINT_01040 (hypothetical protein containing an esterase domain). For the glycoside hydrolases, the four different predicted enzymes are BACINT_01035 (predicted GH43), BACINT_01041 (predicted GH43), BACINT_01042 (predicted GH3), and BACINT_01043 (predicted GH43).

Due to the importance of arabinoxylan for the physiology of the microorganism, as well as the host, the efficient cleavage of ester-linkage side chains is key to the full degradation of the substrate. These enzymes, that likely target esterified arabinoxylan,

present a variable modular organization among their polypeptide sequences (Fig 2.5B). The gene product of BACINT_01033 is composed of a predicted signal peptide and a carbohydrate esterase family 1 module. For BACINT_01035, there is a signal peptide, a GH43 module, and a CBM13 module predicted to bind to xylan. BACINT_01038, a bi-modular protein containing CE6/CE1 does not possess a signal peptide. Moreover, the gene product of BACINT_01038 seemed to be prone to proteases activity by the heterologous expression system, but addition of the protease inhibitor Benzamidine allowed purification of the full construct (Sup. Fig. D.5.). The putative esterase BACINT_01039 possesses a signal peptide and a CE1 family module. Interestingly, BACINT_01040 possesses a signal peptide and an unassigned esterase domain. The product of BACINT_01041 is a single module of GH43 family with no signal peptide. For the BACINT_01042, this protein is composed of a signal peptide and a large module of GH3 family, predicted to act as β-xylosidase. Lastly, BACINT_01043, possesses a signal peptide, a GH43 domain, and a CBM6 predicted to bind arabinoxylan. The products of these genes were recombinantly expressed in *E. coli* and purified close to homogeneity, as depicted by SDS-PAGE analysis (Fig 2.5C). To determine its substrate specificities, each enzyme was incubated with different ferulated oligosaccharides under optimal pH and temperature (Sup. Fig. D.1.). The putative esterases in this cluster demonstrated different esterase specificities (Fig 2.5D). BACINT_01033 has a high feruloyl esterase activity towards FAXX, an arabinoxylan related oligosaccharide, while having little to no activity towards other ferulated oligosaccharides and methyl-ferulate (Sup. Fig. D.2.). The products of BACINT_01038 and BACINT_01039 showed diminished activity towards all the ferulated substrates, including methyl-ferulate. Surprisingly, the hypothetical protein

encoded by BACINT_01040, presents high feruloyl esterase activity towards every ferulated oligosaccharide tested, as well as the synthetic substrate methyl-ferulate. These results demonstrated that BACINT_01033 and BACINT_01040 encode feruloyl esterases, despite the high specificity of BACINT_01033 and promiscuity of BACINT_01040, suggesting a different interaction of both proteins with the substrates. Another relevant activity of carbohydrate esterases, acetyl xylan esterase, was determined by measuring the acetate released from acetylated oat spelt xylan by each putative esterase. Although, BACINT_01033 and BACINT_01040 demonstrate high feruloyl esterase activity, both enzymes have a low ability to remove acetyl groups from the arabinoxylan backbone and *p*NP-acetate. On the other hand, BACINT_01039 and in particular BACINT_01038 have a higher specificity towards acetyl groups decorating the xylan bcakbone, releasing high amounts of acetate from the natural substrate and synthetic substrate (Fig 2.5E and Fig 6S).

Hypothetical domain encoded by Bi1040 is conserved in the bimodular BeGH43/Fae. By sequence analysis of the hypothetical model in Bi1040, we were able to identify a similar domain in the bifunctional enzyme from *Bacteroides* eggerthii BeGH43/Fae, encoded by the HMPREF1016_RS0111555 gene. This hypothetical domain demonstrated high conservation, of ~70% identity with Bi01040, and an alignment between the two different proteins is depicted in Figure 2.6A. Furthermore, the protein encoded by HMPREF1016_RS0111555 presents an interesting modularity, possessing a glycoside hydrolase 43 family in its C-terminus and the hypothetical domain in the N-terminus (Figure 2.6B). This protein was heterologously expressed and purified close to homogeneity to perform biochemical and structural assays (Figure 2.6C). Furthermore,

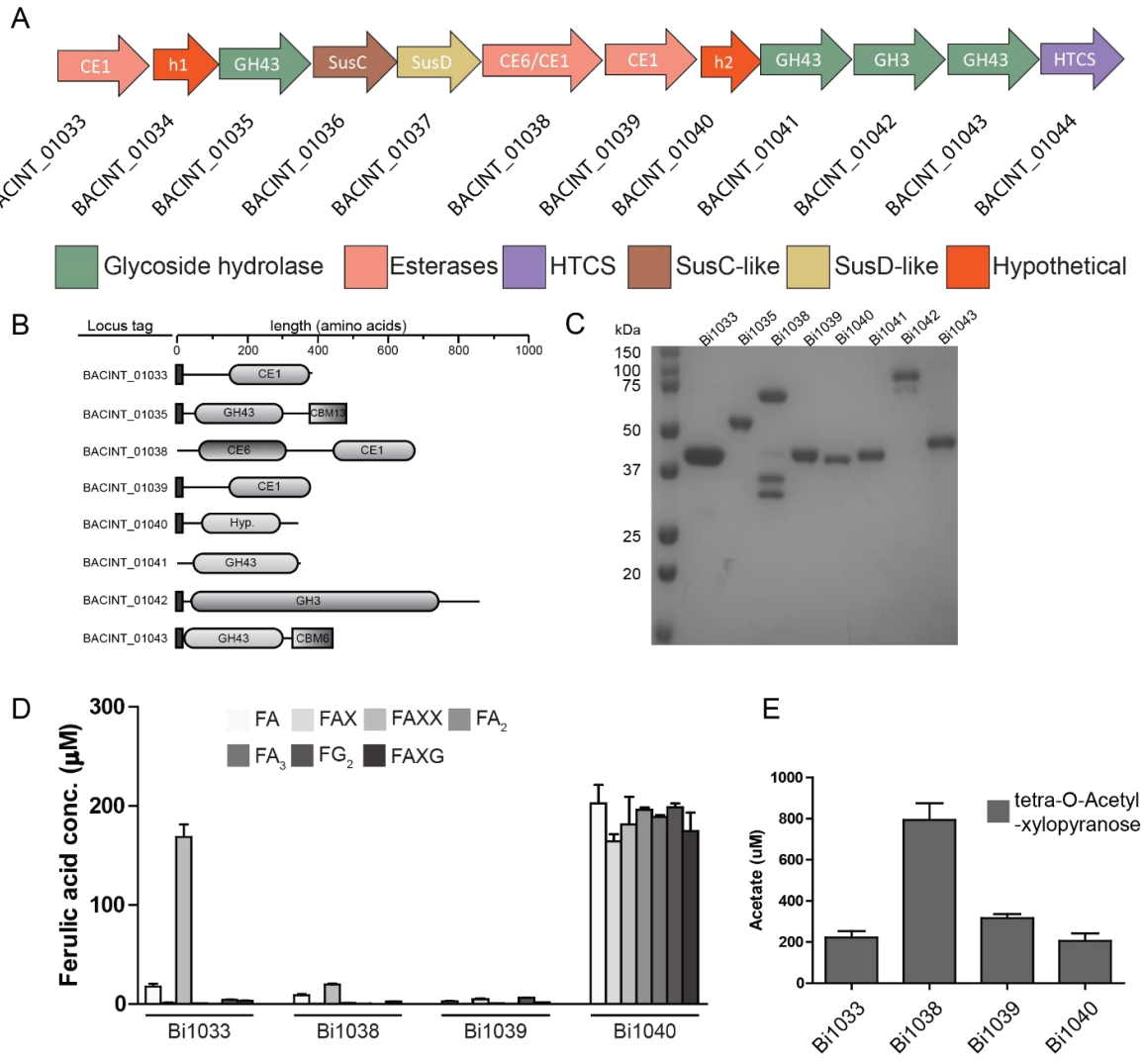


Figure 2.5 - **Esterase-enriched cluster genes demonstrate different esterase specificities.** (A) The genomic context of the esterase-enriched polysaccharide utilization loci. (B) Schematic domain architecture of the putative esterases and putative glycoside hydrolases. (C) SDS-PAGE analysis of the protein expression of putative esterases and glycoside hydrolases. (D) Substrate specificity of the putative esterases in the cluster towards FA, FAX, and FAXX. (E) Acetylxylan esterase activity towards acetylated oat spelt xylan.

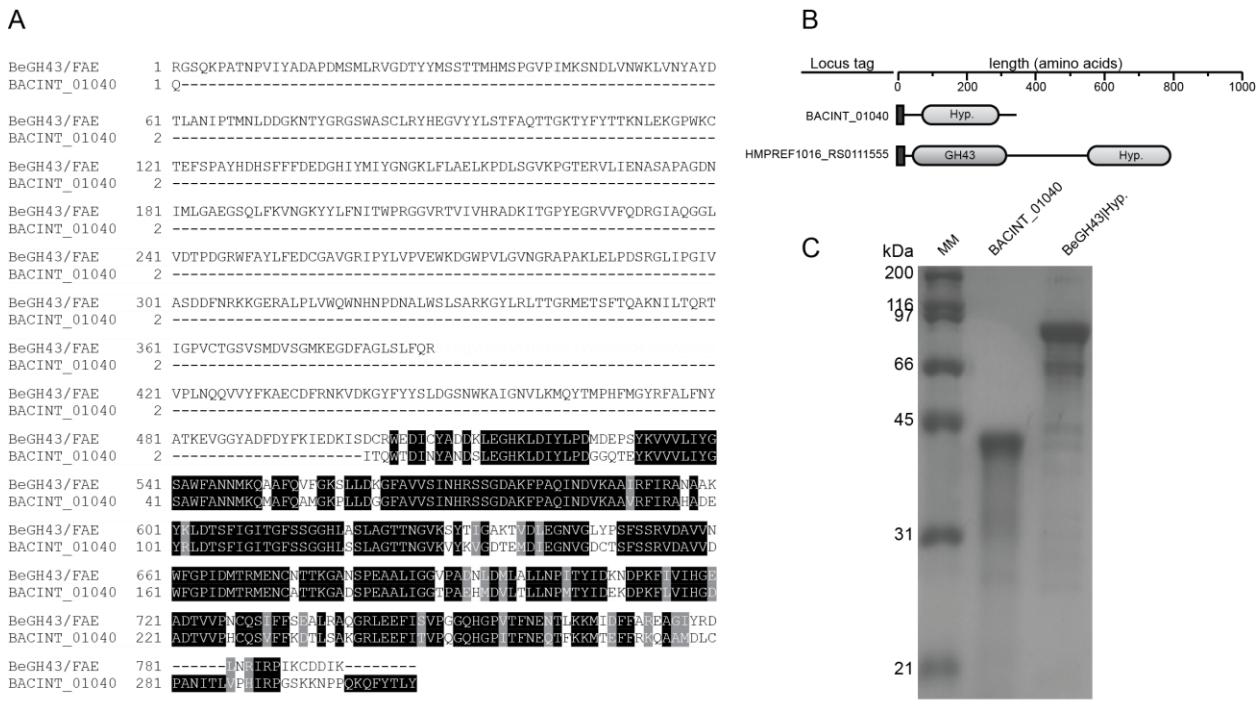


Figure 2.6 - **Comparison of BACINT_01040 hypothetical domain and BeGH43/Fae show high conservation.** (A) Conservation of hypothetical domain encoded by Bi1040 on the N-terminus region of BeGH43/Fae. (B) Protein domain architecture scheme of proteins encoding the hypothetical domain. (C) SDS-PAGE depicting expression and purification of Bi1040 and BeGH43/Fae.

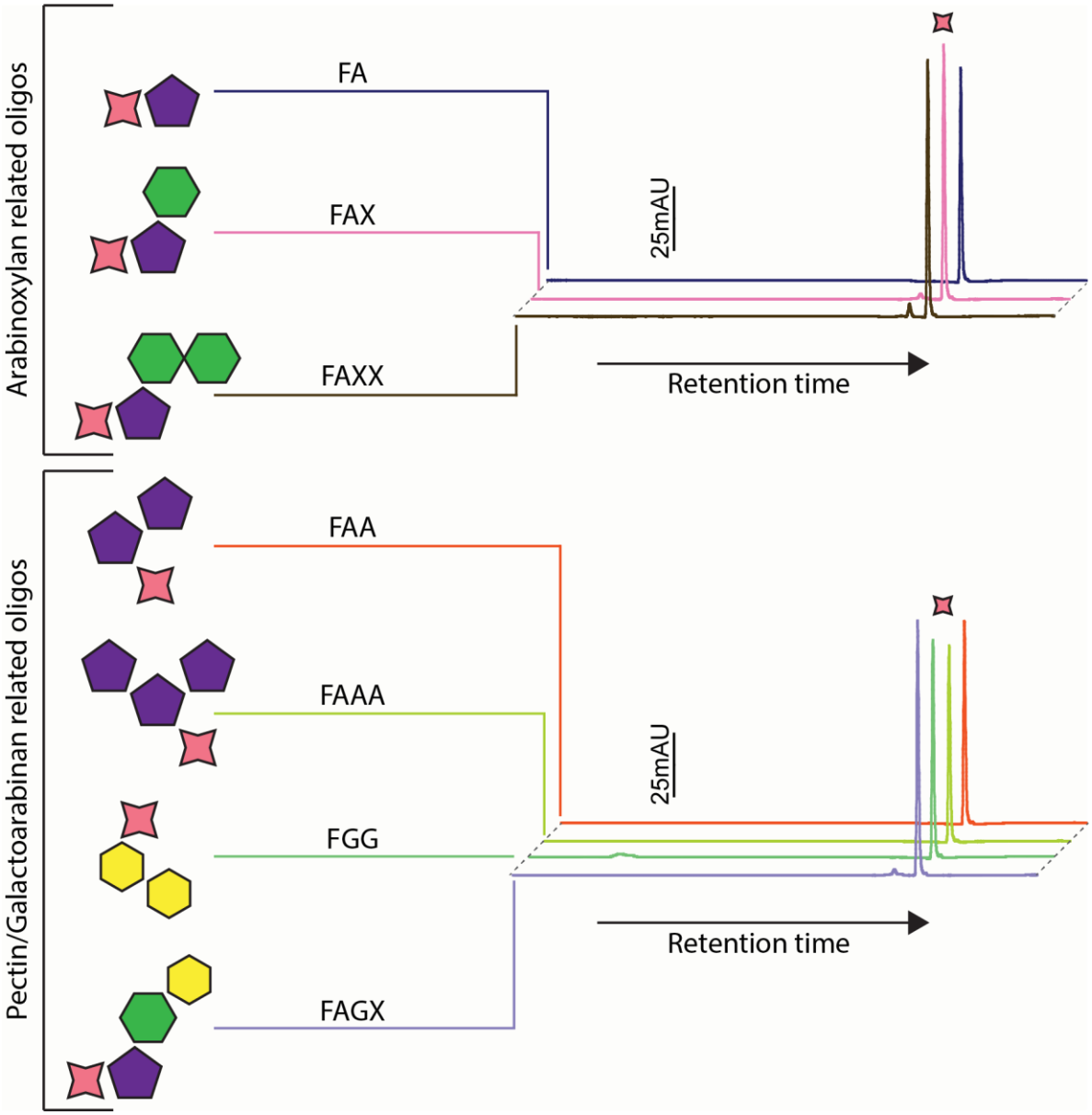


Figure 2.7 - **BeGH43/Fae hypothetical domain has similar substrate specificity and activity to Bi1040.** Activity of the novel bifunctional enzyme encoding hypothetical domain demonstrate broad specificity towards arabinoxylan related ferulated oligosaccharides (FA, FAX, and FAXX) and non arabinoxylan related (FAA, FAAA, FGG, and FAXG) showing full cleavage of ferulic acid side chain.

this hypothetical domain of the BeGH43/Fae demonstrates a very similar activity to the hypothetical domain from Bi1040. This enzyme was incubated for three minutes with seven different ferulated oligosaccharides, including three arabinoxylan related (FA, FAX, and FAXX) and four non-arabinoxylan related substrates (FAA, FAAA, FGG, and FAXG). The enzyme was able to fully remove the ferulic acid side chain from all the substrates (Figure 2.7), similarly to Bi1040, leading to the hypothesis that these two hypothetical domains must work in a very similar mechanism to cleave ferulic acid. Furthermore, the GH43 module encoded by this bifunctional enzyme demonstrates an arabinoxylan dependent arabinofuranosidase activity (Sup. Fig. D.6.).

Hypothetical domain encoded by Bi1040 and BeGH43/Fae may be part of a new family of carbohydrate esterase. The two hypothetical domains characterized in this work demonstrated a promiscuous activity towards different ferulated oligosaccharides. Due to lack of studies demonstrating such substrate specificity in feruloyl esterases, we sought to understand if this domain might be part of a new carbohydrate esterase family encoded by the phylum Bacteroidetes. We collected the carbohydrate esterases sequences of the polysaccharide utilization database (PULDB) to understand the relationship of the hypothetical domains with the currently annotated carbohydrate esterases in Bacteroidetes. By sequence similarity, we generated a tree based on the polypeptide sequence of each different family to identify how the hypothetical domains clustered. In Figure 2.8, it is demonstrated that the hypothetical domains do not cluster with any known family encoded in the different Bacteroidetes species. This result, combined with the wide substrate specificity towards ferulated

substrates demonstrated by the hypothetical domain, further strengthens the hypothesis that this is a new carbohydrate esterase family.

Structural characterization of the acetyl xylan esterase Bi1039. The native structure of Bi1039 was solved and characterized as an asymmetrical homodimer (Figure 2.9A). Furthermore, overlapping of both chains indicates a structural conservation in each monomer (Figure 2.9B). The catalytic triad of esterases is commonly composed of aspartate (glutamate), histidine, and serine. As we solved the structure in the apo form, we superimposed the acetyl xylan esterase encoded by Bi1039 with a previously characterized esterase (PDB:1JT2) to better identify the catalytic domain. Overlapping these two esterases revealed conservation of the catalytic amino acids histidine and serine, while the aspartate present in 1JT2 is a glutamate in Bi1039 (Sup. Fig. D.3.). Thus, we hypothesize that the mechanism employed by Bi1039 follows the canonical catalytic triad presented by most studied carbohydrate esterases. The de-protonated glutamate carboxylic acid transfers an electron to the histidine imidazole ring, thereby allowing the displacement of the sugar backbone and acylation of the serine. Following serine de-acylation by an activated water molecule, the water molecule attacks the ester linkage of the acyl-serine (see Fig 1.3). The catalytic triad is depicted in figure 2.9C.

Substrate specificity of the glycoside hydrolases encoded in the esterase-enriched cluster towards different xylan substrates. Hydrolytic activities of the glycoside hydrolases were studied by incubating each xylan substrate with each enzyme. All four enzymes demonstrated activity towards xylan substrates, with the highest activity towards the different arabinoxylans (Fig 2.10A). Based on these results, we determined the activity of BACINT_01035 and BACINT_01041 as endoxylanases, releasing large

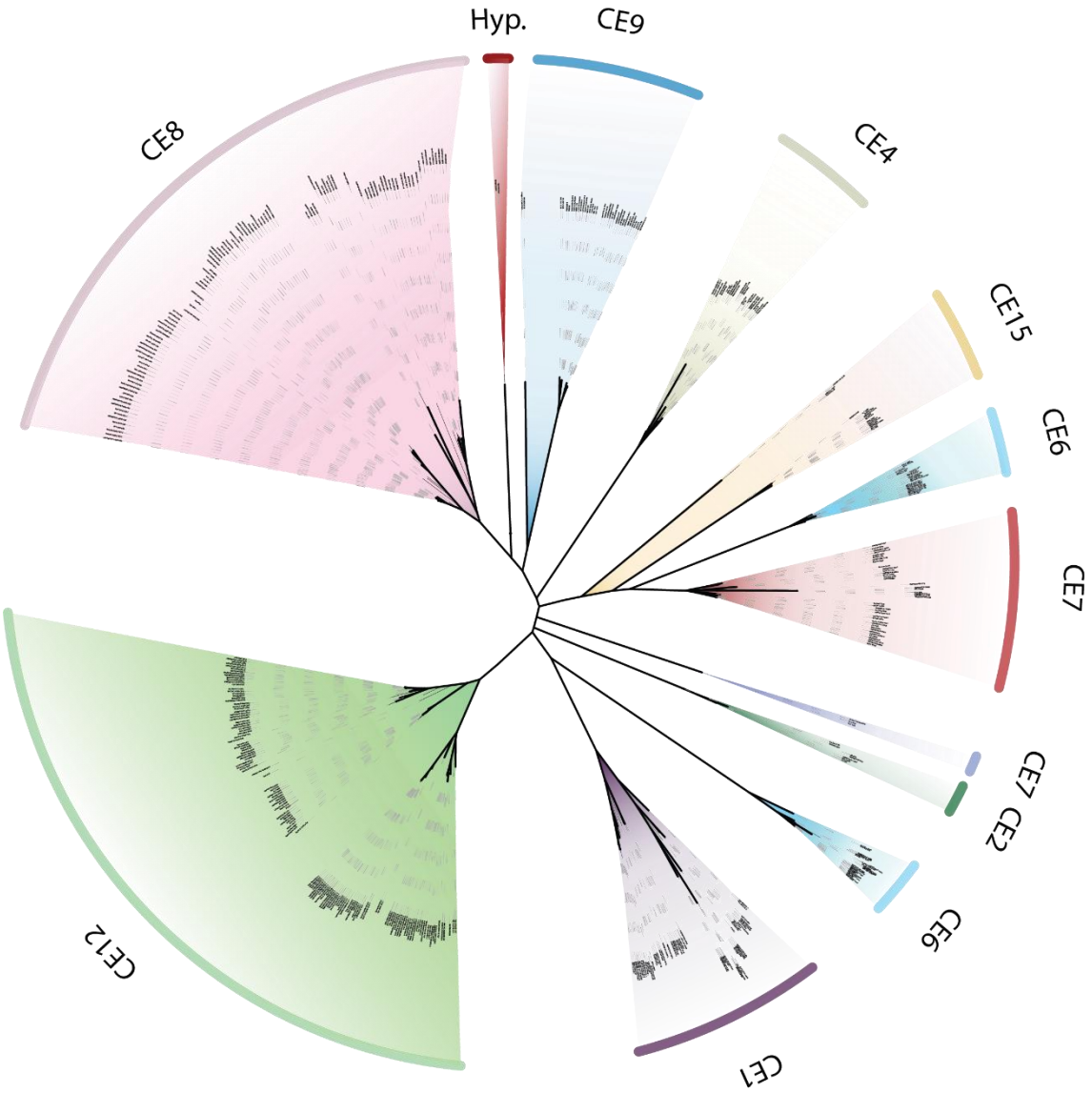


Figure 2.8 - **Hypothetical domain does not cluster with other carbohydrate esterases families.** Carbohydrate esterase encoded by Bacteroidetes species phylogeny shows that hypothetical domains (red) do not cluster with known families represented in Bacteroidetes.

amounts of reducing sugars. Further analysis of BACINT_01042 towards xylooligosaccharides demonstrated a β-xylosidase activity by cleavage of the different oligosaccharides to their monomeric sugars (Fig 2.10B).

Synergistic activity of carbohydrate esterases and glycoside hydrolases encoded in the esterase-enriched cluster increases the release of ferulic acid and reducing sugars. To determine the functional interactions of the carbohydrate esterases and glycoside hydrolases encoded by the esterase-enriched cluster, we incubated single and mixtures of enzymes to measure the ferulic acid and sugar reducing ends released. Combination of the endoxylanases (BACINT_01035 and BACINT_01041) with the ferulyol esterases (BACINT_01033 and BACINT_01040) resulted in a synergistic effect in both releasing ferulic acid (< 2-fold) and reducing ends (~2-fold) in destarched wheat bran (Fig 2.11A) and insoluble wheat arabinoxylan (Fig 2.11B). Further combinations, including the accessory β-xylosidase and BACINT_01043, increased the amounts of end products released from the reaction with each substrate. A mixture of all the enzymes encoded by the cluster reached reducing sugar levels up to 8mM for both substrates, and ~50 uM and 60 uM of ferulic acid from destarched wheat bran and insoluble wheat arabinoxylan, respectively. These results demonstrated the synergism of the cluster towards degrading ferulated arabinoxylan, allowing an efficient depolymerization by the human colonic bacterium *B. intestinalis*.

Esterases encoded by the esterase-enriched cluster works synergistically with homologous and heterologous endoxylanases. In addition to the synergistic effect of esterases and glycoside hydrolases within the esterase-enriched cluster, we determined the synergistic effect of homologous and heterologous endoxylanases with the feruloyl

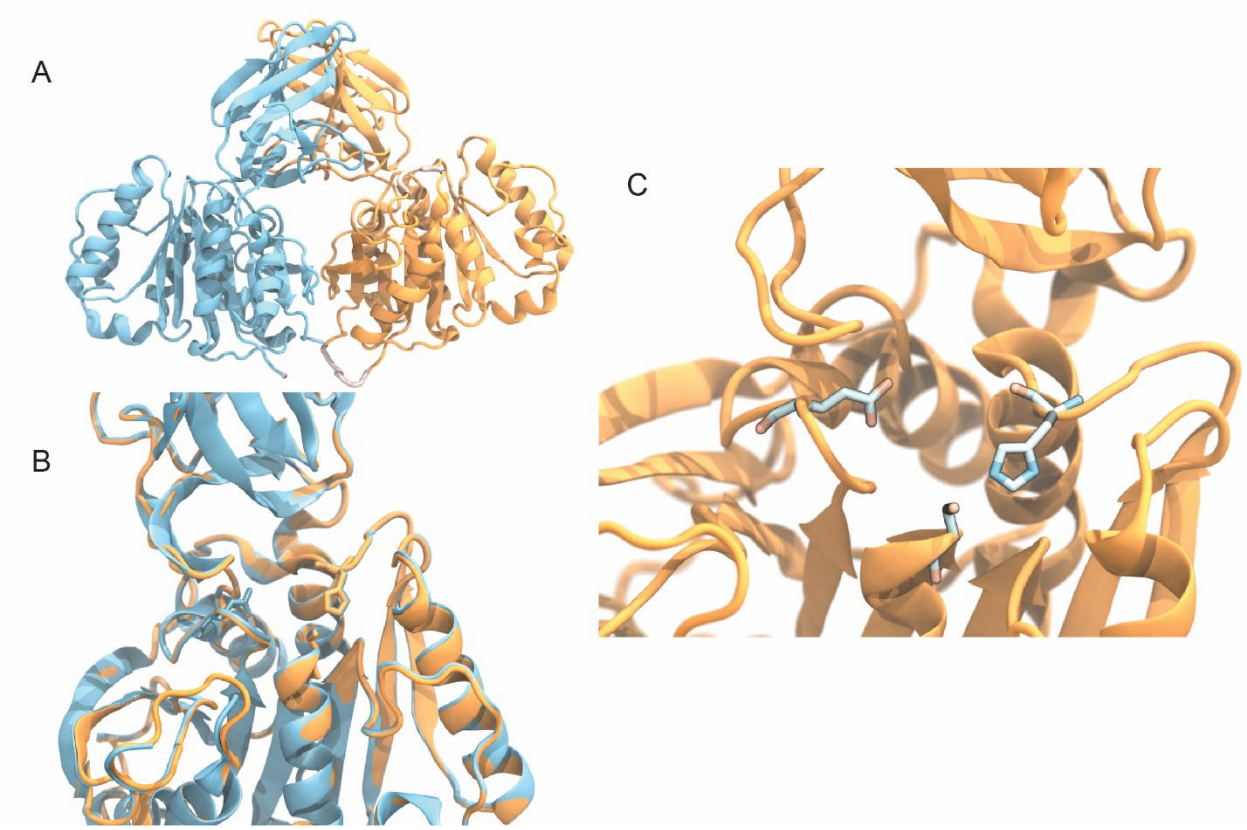


Figure 2.9 - **Acetyl xylan esterase encoded by esterase enriched cluster forms a homodimer.** Chain A and Chain B forming a homodimer (A), and overlapping representation of both chains (B). A catalytic triad composed of glutamate, histidine and serine coordinates the enzyme activity (C).

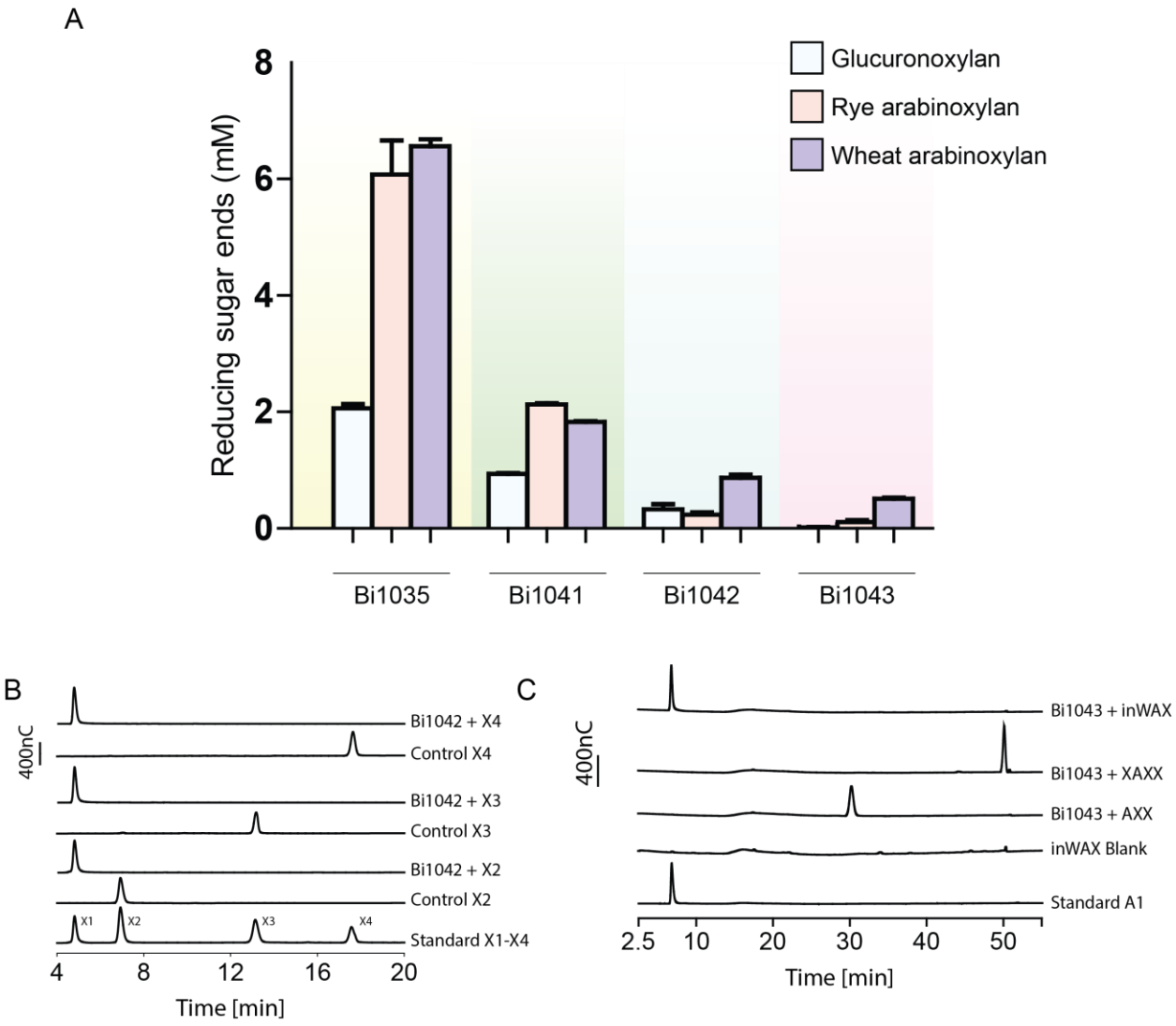


Figure 2.10 - **Hydrolytic activities of the glycoside hydrolases encoded in the esterase-enriched cluster.** (A) Reducing sugar released of xylan substrates by putative glycoside hydrolases. (B) β-xylosidase activity of BACINT_01042 towards xylo-oligosaccharides. (C) Arabinoxylan dependent α-arabinofuranosidase activity of BACINT_01043.

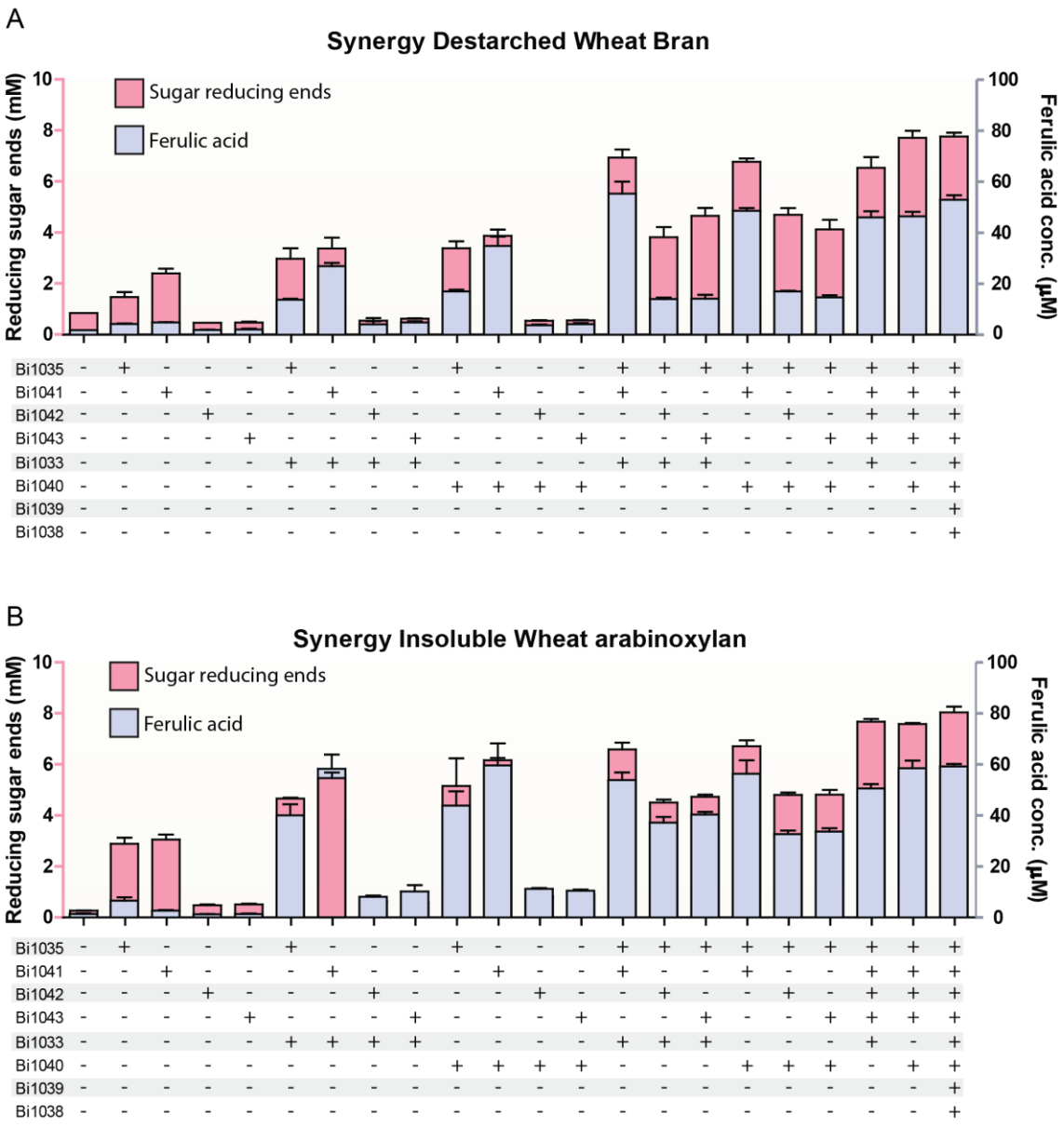


Figure 2.11 - **Synergistic activity of enzymes encoded by esterase-enriched cluster increase sugar and ferulic acid release.** The synergistic activity of the enzymes encoded by the esterase-enriched cluster were assessed by incubating the enzymes with destarched wheat bran (A) and insoluble WAX (B). Released reducing sugars were measured by *p*HBAH assay using known concentrations of xylose as standard and released ferulic acid was measured by HPLC-PAED using known concentrations of ferulic acid as standard.

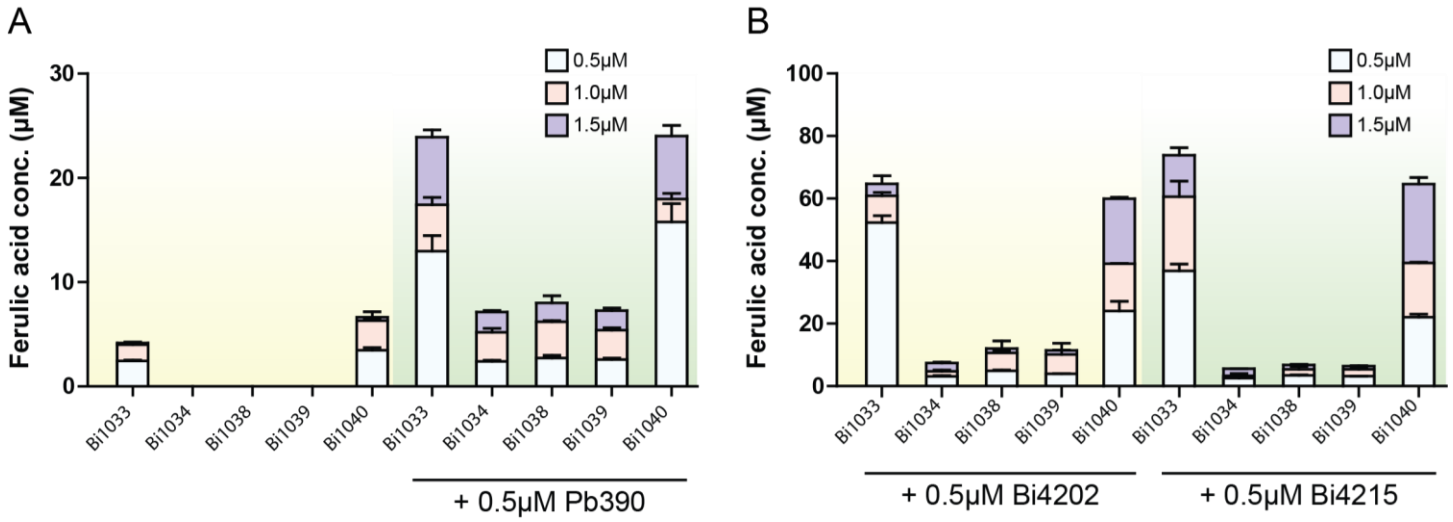


Figure 2.12 - **Synergistic activity of putative esterases from the esterase-enriched cluster with homologous and heterologous endoxylanases.** The solo and synergistic activity of the different concentration of enzymes encoded by the esterase-enriched cluster and homologous or heterologous enzymes were assessed by incubating the enzymes with insoluble WAX. (A) Single enzymes and combination with heterologous endoxylanase. (B) Synergistic activity of esterases with two previously characterized endoxylanases from *B. intestinalis*. Released ferulic acid was measured by HPLC-PAED using known concentrations of ferulic acid as standard.

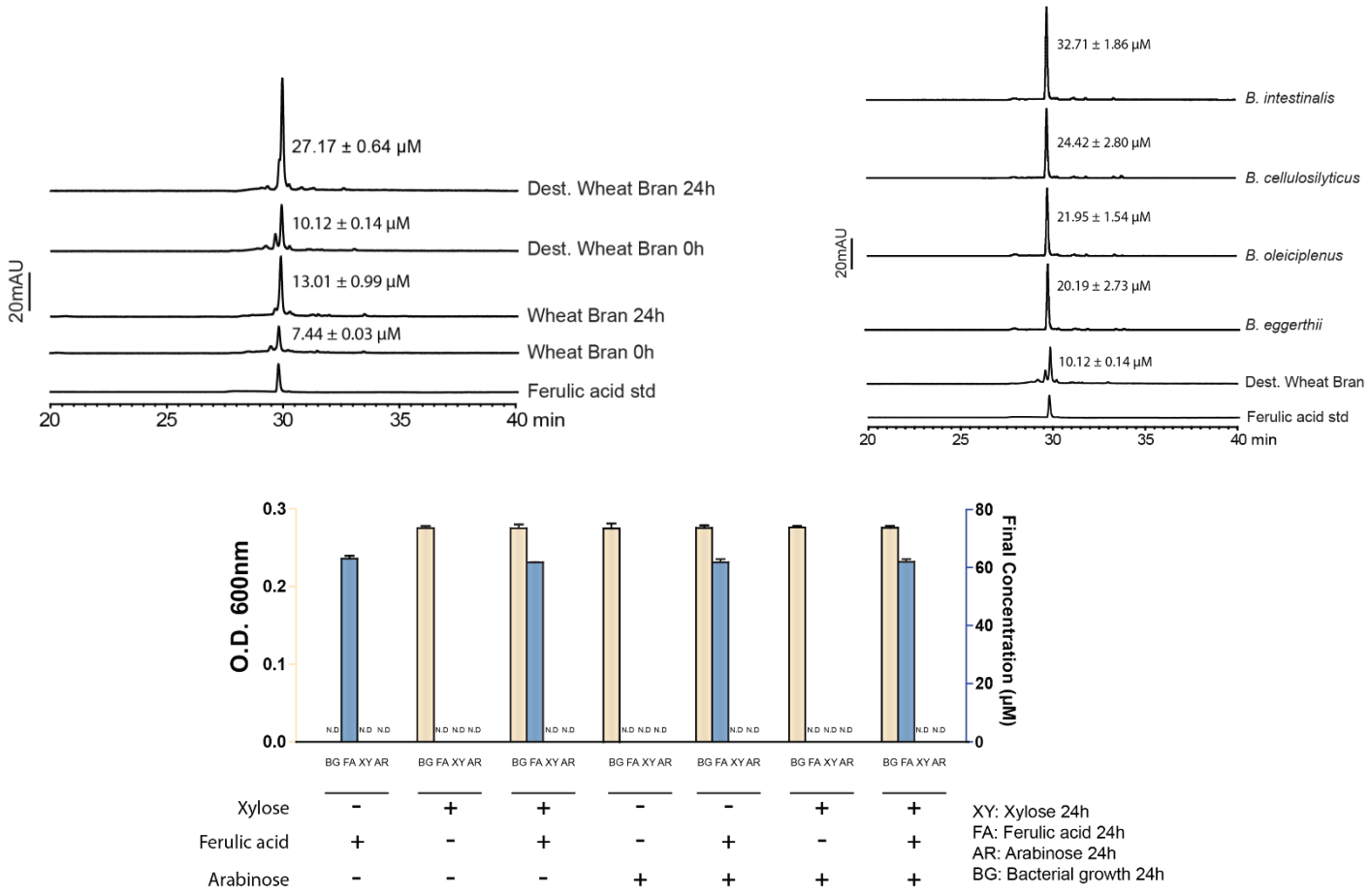


Figure 2.13 - **Accumulation of ferulic acid in the spent medium of *B. intestinalis* indicates a lack of ferulic acid utilization.** (A) HPLC-PAED chromatogram of *B. intestinalis* spent medium using wheat bran and destarched wheat bran as solo carbon source, demonstrating increasing release of ferulic acid. (B) Growth of *B. intestinalis* in minimal medium containing ferulic acid as sole carbon source and supplemented with different arabinoxylan monosaccharides.

esterases. The combination of each previously characterized endoxylanases from *Prevotella bryanti* (Pb390) and *B. intestinalis* (Bi4202 and Bi4215) with different concentrations of each esterase demonstrated a large synergistic effect in the release of ferulic acid, especially for the feruloyl esterases. The levels of ferulic acid released by BACINT_01033 combined with Pb390, Bi4202, and Bi4215 increased <5-fold, <12-fold, and 14-fold, respectively, compared to the enzyme alone towards insoluble WAX. For the synergy of BACINT_01040 combined with the endoxylanases, there was an increase of ferulic acid released of <3-fold, 7.5-fold, and 8-fold, when in combination with Pb390, Bi4202, and Bi4215, respectively (Fig 2.12). These results underscore the ability of the esterases to work synergistically with enzymes responsible for degrading the arabinoxylan backbone.

***B. intestinalis* does not metabolize ferulic acid.** *B. intestinalis* growth on the complex natural substrates wheat bran and destarched wheat bran demonstrated the ability of this bacterium to remove and accumulate ferulic acid side chain, 2-fold and 3-fold, respectively, from arabinoxylan backbone (Fig 2.13A). Furthermore, we demonstrate the ability of *B. intestinalis* to generate different short-chain fatty acids, increasing its protein content during growth in wheat bran and destarched wheat bran (Figure S3). To further understand the ability of different *Bacteroides* species to accumulate ferulic acid after growth in a highly esterified substrate, we grew two other *Bacteroides* species that encode the esterase-enriched cluster (*B. cellulosilyticus* and *B. oleiciplenus*) and *B. eggerthii*, that despite not encoding this operon seems to have a mechanism to release ferulic acid from esterified substrates as demonstrated by the activity of BeGH43/Fae. Similarly to *B. intestinalis*, the three bacteria seem to be able to grow on the complex

arabinoxylan substrate, destarched wheat bran, while accumulating ferulic acid in the spent medium (Figure 2.13B). To determine the ability of *B. intestinalis* to consume and/or metabolize the ferulic acid released from arabinoxylan, we tested the growth of this bacterium under several different conditions. Initially, ferulic acid did not support growth of this bacterium as a sole carbon source. Moreover, growth on limiting amounts of different monosaccharides (xylose, arabinose, and arabinose:xylose) with or without ferulic did not lead to consumption or metabolism of ferulic acid (Fig 2.13C). Therefore, the free ferulic acid supplemented in the medium did not show any improvements in growth or metabolism of *B. intestinalis*. These results demonstrate the ability of the colonic bacterium, *B. intestinalis*, to cleave ferulic acid side chains in order to access the polysaccharide backbone for energy, while accumulating the natural phenolic compound in the medium. This accumulation of ferulic acid, combined with previous work related to the benefits of ferulic acid for the host, leads to a potential beneficial interaction between *B. intestinalis* and the human host.

The colonic bacterium, *B. intestinalis*, modulates immune response of human intestinal cells. Based on previous research about the therapeutic potential of ferulic acid, we determined the effect of the ferulic acid released by *B. intestinalis* and human cells. To identify the interaction of released ferulic acid present in the spent medium of *B. intestinalis* grown in wheat bran and destarched wheat bran, and human colonocytes, we measured the production of cytokines (IL-4, IL-10, IL-12, IFN-α, and IFN-β) by ELISA and RT-qPCR. The cells treated with *B. intestinalis* spent medium demonstrated a significantly enhanced expression of each cytokine compared to the controls (Fig 2.14A-E), while the positive control (10-100uM ferulic acid) increased the

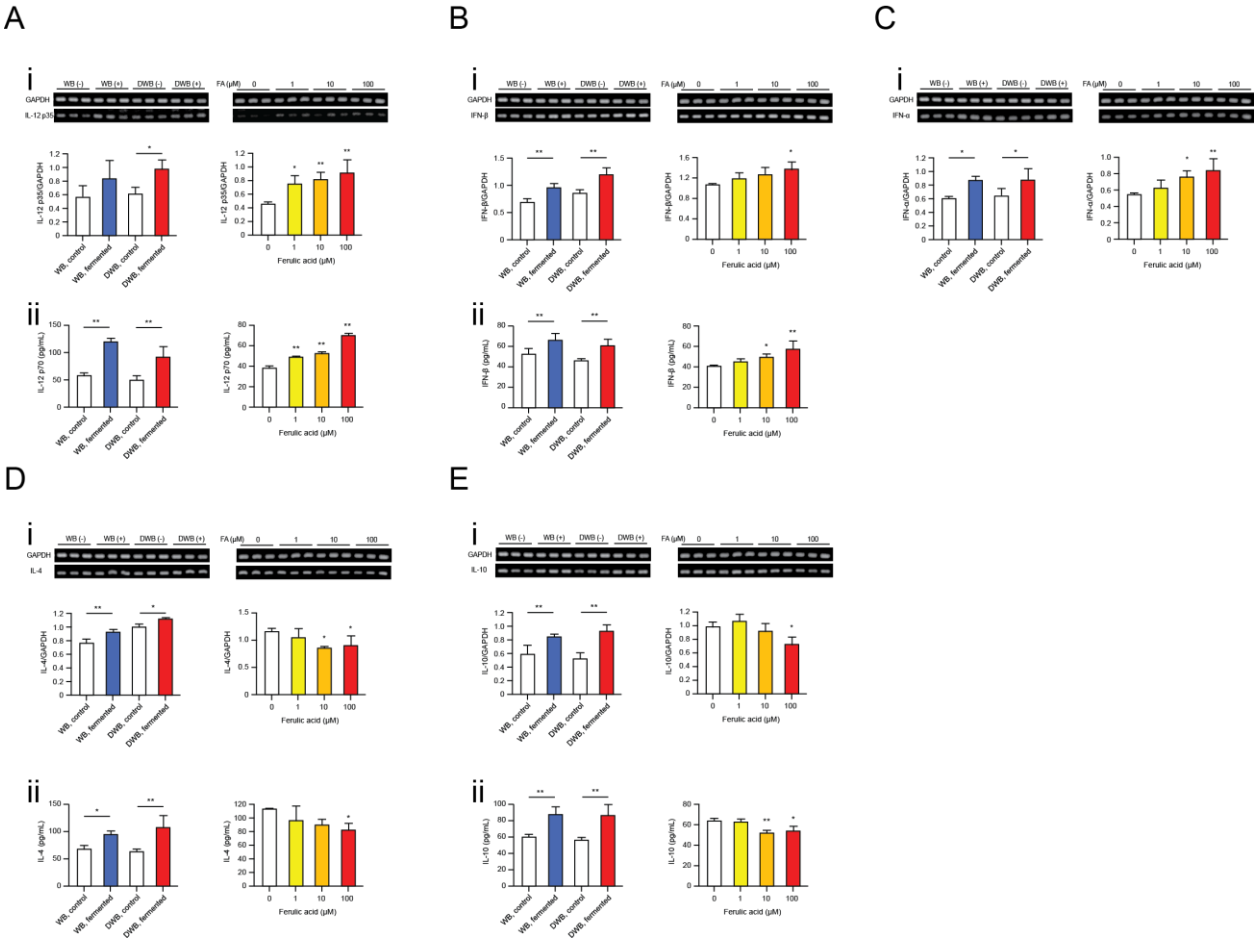


Figure 2.14 - ***Bacteroides intestinalis*** **spent medium modulates cytokine expression in Caco-2 cells.** Spent medium of *B. intestinalis* grown in Wheat bran (WB, blue) and Destarched wheat bran (DWB, red) increased the production of IL-12, IFN-β, IFN-α, IL-4, and IL-10 demonstrated through RT-qPCR (Ai, Bi, Ci, Di, and Ei respectively) and ELISA (Aii, Bii, Dii, and Eii respectively). Ferulic acid in three concentrations (1, 10, and 100μM) was used as positive control, although the increased production of IL-4 and IL-10 can be attributed to ferulic acid released by *B. intestinalis*. These results were obtained in collaboration with Professor Esteban Gabazza and Professor Corina D'Alessandro-Gabazza from Mie University (Japan).

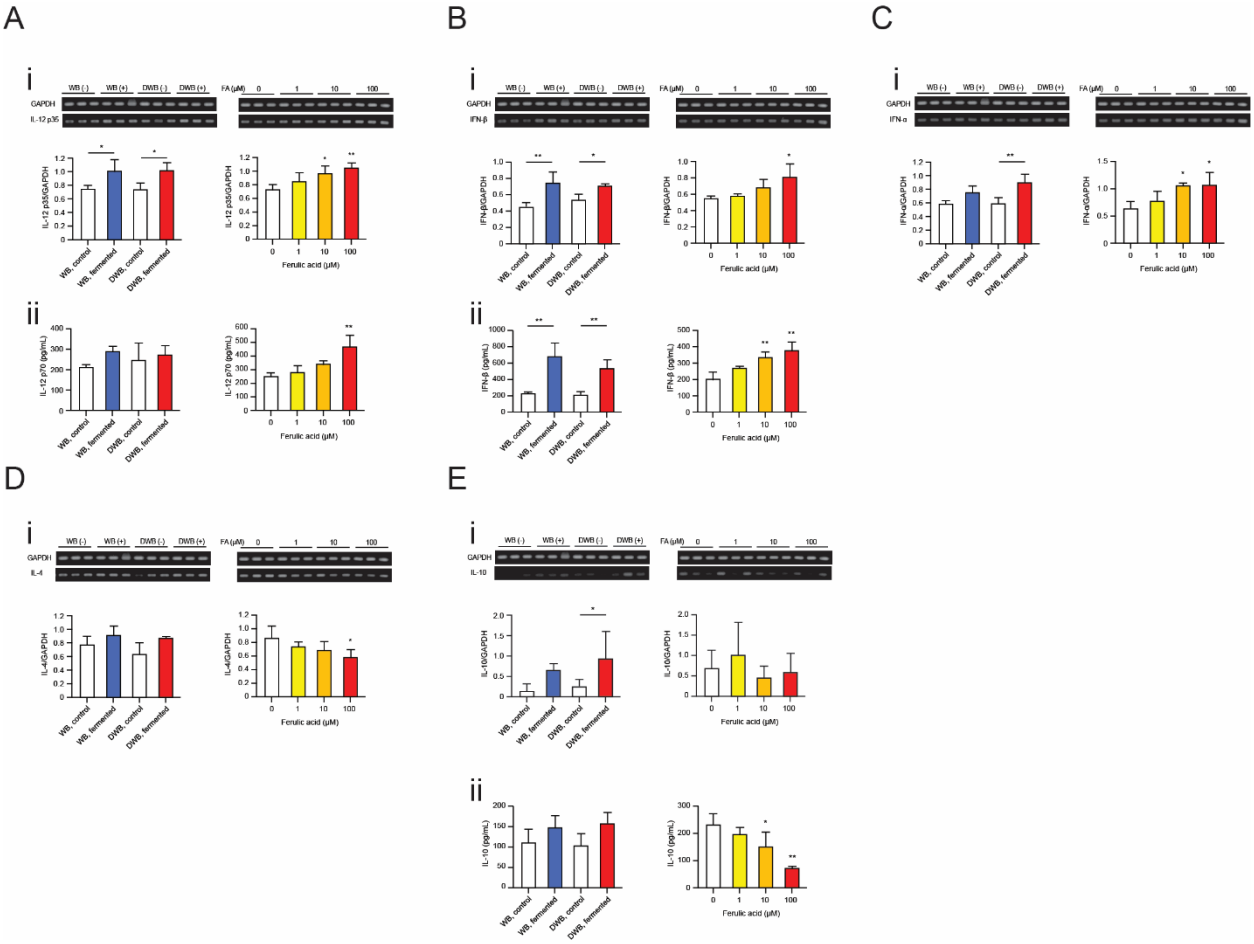


Figure 2.15 - ***Bacteroides intestinalis*** **spent medium modulates cytokine expression in HIEC-6 cells.** Spent medium of *B. intestinalis* grown in Wheat bran (WB, blue) and Destarched wheat bran (DWB, red) increased the production of IL-12, IFN-β, IFN-α, IL-4, and IL-10 demonstrated through RT-qPCR (Ai, Bi, Ci, Di, and Ei respectively) and ELISA (Aii, Bii, and Eii respectively). Ferulic acid in three concentrations (1, 10, and 100μM) was used as positive control, although the increased production of IL-4 and IL-10 cannot be attributed to ferulic acid released by *B. intestinalis.* These results were obtained in collaboration with Professor Esteban Gabazza and Professor Corina D'Alessandro-Gabazza from Mie University (Japan).

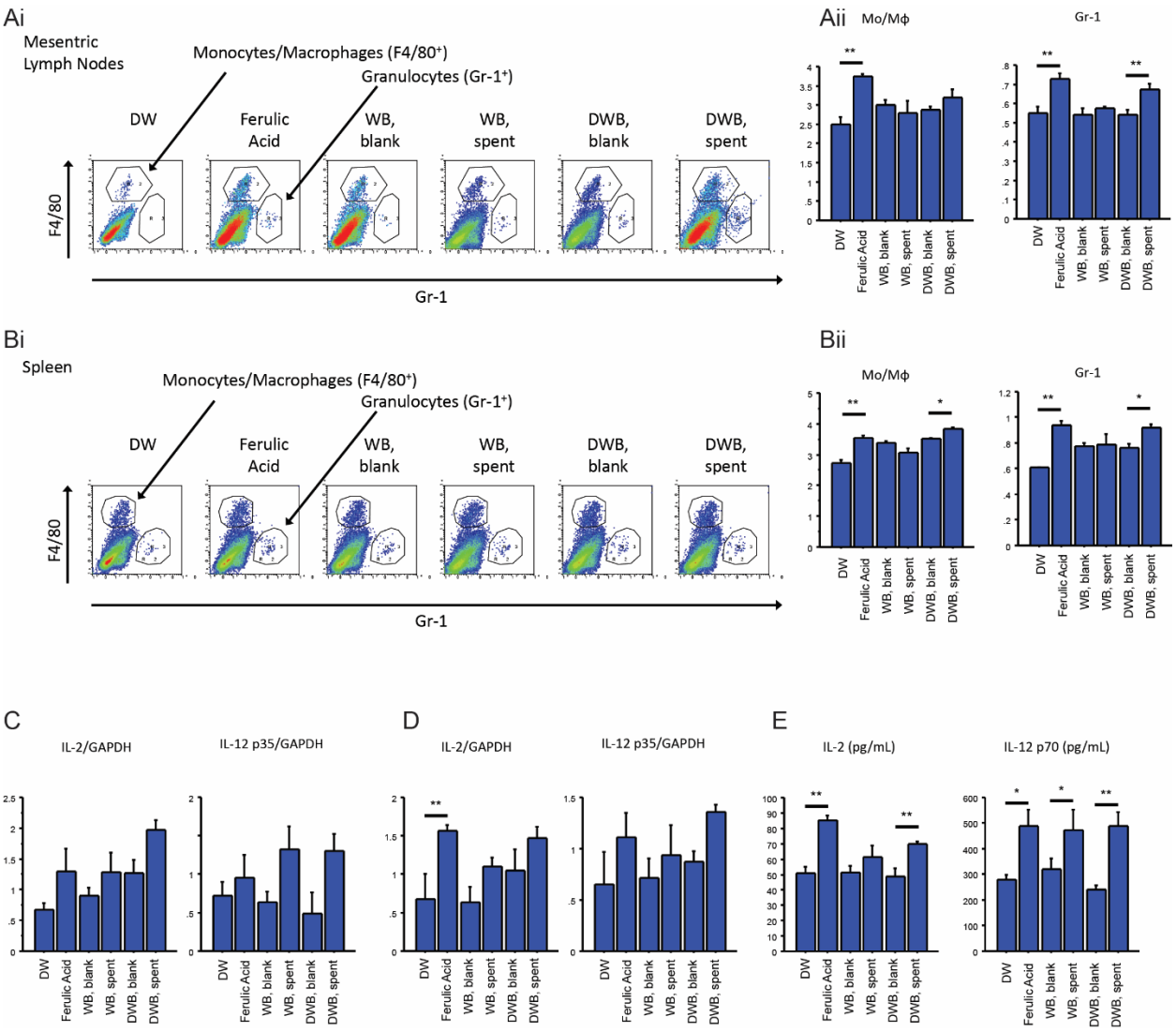


Figure 2.16 - ***Bacteroides intestinalis*** **spent medium modulates activation of granulocytes and macrophages associated with Th1 cytokines expression.** Flow cytometry analyses demonstrate increased numbers of granulocytes and macrophages in mesenteric lymph nodes (Ai and Aii) and spleen (Bi and Bii). Increased expression of Th1 cytokines IL-2 and IL-3 measured by RT-qPCR in the mesenteric lymph nodes (C) and spleen (D), as well as measured in plasma by ELISA (E). These results were obtained in collaboration with Professor Esteban Gabazza and Professor Corina D'Alessandro-Gabazza from Mie University (Japan).

expression of IL-12, IFN-α, and IFN-β (Fig 2.14A-C). Moreover, the positive control decreased the expression of IL-4 and IL-10 (Fig 2.14D-E). In a similar way, the spent medium of *B. intestinalis* grown in wheat bran and destarched wheat bran is able to elicit a response of human intestinal epithelium cells. The positive control showed an increase in the production of pro-inflammatory cytokines IL-12, IFN-β, and IFN-α (Fig 2.15A-C), while decreasing the expression of anti-inflammatory IL-4 and IL-10 (Fig 2.15D-E)The HIEC-6 cells incubated with the spent medium demonstrated an increase production of the different cytokines IL-12, IFN-β, IFN-α, IL-4 and IL-10 (Fig 2.15A-E).

Ferulic acid released by the esterase-enriched cluster is able to increase production of Th1-associated cytokines. Although human tissue culture provides evidence of modulation by *B. intestinalis* spent medium, the complexity of an in vivo experiment is still not achievable with these. To determine the influence of *B. intestinalis* spent medium towards a mouse model, we gavaged the mouse with spent medium for seven days and verified the cell expression and cytokine expression in the spleen and mesenteric lymph nodes. The treatments of spent medium and positive control (ferulic acid) demonstrate an elevation of macrophages (Figure 2.16Ai and Aii) and granulocytes (Figure 2.16Bi and Bii) related to adaptive immune system stimulation. Moreover, there was also an increased expression of IL-2 and IL-12 in both the spleen (Figure 2.16C) and lymph nodes (Figure 2.16E).

2.3 Discussion

One of the most abundant groups of microorganisms in the human gastrointestinal tract, the *Bacteroides* phylum, possesses an outstanding array of genes targeting dietary

and host polysaccharides (El Kaoutari, Armougom et al. 2013). These organized polysaccharide-targeting genes are key to successful competition and colonization in the human gut (Xu, Mahowald et al. 2007). The polysaccharide utilization loci are usually composed of oligosaccharide binding and transport membrane associated proteins, glycoside hydrolases responsible for cleaving the polysaccharides, and regulators that allow increased transcription of the operon (Martens, Chiang et al. 2008, Mackie and Cann 2018). Therefore, these PULs confer a remarkable competitive advantage, in the presence of the substrate in the host's diet, due to its tight and fine-tuned regulation. Furthermore, the Bacteroidetes possess one of the largest and most diversified repertoire of glycoside hydrolases and polysaccharide lyases, especially *Bacteroides intestinalis* and *Bacteroides cellulosilyticus (El Kaoutari, Armougom et al. 2013, McNulty, Wu et al. 2013)*.

Recent studies have demonstrated the importance of diet-related arabinoxylan, such as whole cereal grains, to normal gut function and health (Smith and Tucker 2011). Furthermore, whole cereal associated polysaccharides have been extensively studied as a prebiotic dietary supplement (Duncan, Holtrop et al. 2004). Thus, understanding the degradation and fermentation of arabinoxylan, one of the major polysaccharides from whole cereals, such as wheat and oats, is key to allowing dietary interventions with precision. In fact, recent reports have demonstrated that gut bacteria, such as B. *ovatus* and *B. intestinalis*, encodes one of the largest xylan-degrading enzymatic repertoire. For instance, *B. intestinalis* grown in wheat arabinoxylan up-regulates nearly 70 genes, when compared to growth on its constituent monosaccharides xylose and arabinose, with 22 of these genes encoded in a single PUL (Zhang, Chekan et al. 2014, Wang, Pereira et al.

2016). Studies have demonstrated the ability of this bacterium to fully depolymerize arabinoxylan into fermentable sugars, which generate short-chain fatty acids used by the host as a source of energy (Wang, Pereira et al. 2016). Interestingly, growth of three different *Bacteroides* species (*Bacteroides intestinalis*, *Bacteroides cellulosyliticus*, and *Bacteroides oleiciplenus*) in a more esterified arabinoxylan substrate demonstrate up-regulation of a different arabinoxylan degrading PUL compared to the constituent monosaccharides. These esterase-enriched PULs encode several esterases, responsible for cleaving decorating moieties such as ferulic acid and acetyl groups from the arabinoxylan backbone. In fact, *B. intestinalis*, *B. cellulosilyticus* and *B. oleiciplenus* encode four, three, and one esterase in this single PUL respectively. Moreover, despite genetic shifts, these PUL possess all the core genes for full degradation and transport of target polysaccharides, including the *susC/susD* homologs, glycoside hydrolase-encoding genes, and their associated hybrid two-component systems. Remarkably, high expression of this cluster is only stimulated by highly esterified substrates, in particular ferulated substrates such as arabinoxylan-oligosaccharides derivatives.

The diversity of esterases in this cluster, combined with the conservation of these proteins in these different *Bacteroides* species, alludes to the importance of complex arabinoxylan substrates in the host diet. The putative esterases in this cluster demonstrate different activities and substrate specificities. The proteins encoded by Bi1038 and Bi1039 act as acetyl-xylan esterases, cleaving the acetyl groups from both synthetic substrates, as well as complex acetylated oat spelt xylan. Furthermore, the putative feruloyl esterase encoded by Bi1033 is able to cleave ferulic acid side chain from large ferulated arabinoxylan oligosaccharides and insoluble wheat arabinoxylan.

Remarkably, the hypothetical enzyme encoded by Bi1040 is the most prolific and promiscuous feruloyl esterase characterized to date. This enzyme possesses a rapid feruloyl esterase activity towards several different ferulated substrates, including the synthetic substrate methyl-ferulate and the arabinoxylan related oligosaccharides (FA, FAX, and FAXX) and non-arabinoxylan related oligosaccharides (FAA, FAAA, FGG, and FAXG). This novel enzyme is able to fully cleave the ferulic acid side chain of these substrates, allowing access of the glycoside hydrolases to the glycosidic linkages.

Interestingly, the repertoire of glycoside hydrolases in this cluster is composed of three different GH43 families and a GH3 family. Glycoside hydrolase 43 families are known as some of the most diversified activities in CAZy enzymes, ranging from β-xylosidases, α-arabinofuranosidases, to endoxylanases. Despite the narrow diversity in GH families encoded by the esterase-enriched cluster, this cluster possess the entire range of activities to fully depolymerize the arabinoxylan backbone into its fermentable monosaccharides. The proteins encoded by Bi1035 and Bi1041 demonstrate endoxylanase activity, releasing large amounts of reducing ends from polysaccharides, while Bi1042 is able to fully degrade xylooligossacharides into their unit sugars, xylose. Interestingly, Bi1043 seems to act as an arabinoxylan dependent α-arabinofuranosidase. This enzyme exhibits activity only towards the large polysaccharide. We hypothesized that this activity depends on the CBM6 binding to the xylan backbone to allow protein stabilization for its activity. In fact, the enzymes encoded by this cluster act in synergism towards different complex esterified arabinoxylan substrates, such as insoluble wheat arabinoxylan and wheat bran, highlighting the physiological activity of this cluster for the bacteria.

Due to the impressive activity of the novel esterase encoded by Bi1040, we identified a bifunctional enzyme, from *Bacteroides eggerthii*, that encodes a GH43 family enzyme combined with a conserved hypothetical domain. Similarly, this enzyme demonstrates a wide promiscuity towards ferulated substrates.

Fascinatingly, the bacteria studied in this work seem to employ the esterase-enriched cluster to remove the ferulic acid side chain only in order to access the arabinoxylan backbone for energy. Culturing of these bacteria in highly esterified substrates leads to accumulation of ferulic acid. Thus, our results demonstrate that primary degraders of arabinoxylan release ferulic acid into the human colon. This ferulic acid pool should be available for either absorption by the gut epithelial cells or further metabolized by the resident gut microbiota into other phenolic derivatives that can be absorbed by the host.

To address the potential of ferulic acid released by *Bacteroides* growth in complex esterified arabinoxylan substrates, i.e., wheat bran and de-starched wheat bran, we used two different human gastrointestinal cell lines HIEC-6 and Caco-2 cells. In fact, the spent medium and the positive control were able to increase the expression of important Th1 related cytokines such as IL-12, IFN-α, and IFN-β compared to the un-inoculated medium. These cytokines play an important role in the adaptive immune system and its response to intracellular pathogens and viral infections, respectively. Interestingly, the anti-inflammatory cytokines IL-4 and IL-10, which play an active role in differentiation of macrophages into wound repair W2 cells and regulation of anti-inflammatory processes correspondingly, did not demonstrate increased expression in the positive control ferulic acid treatment. Despite the lack of effect by the positive control, the spent medium

treatments were able to significantly increase the expression of these key cytokines. The current hypothesis is that short-chain fatty acid, such as acetate, produced by fermentation by the bacteria growth (Figure 2.12A) is able to modulate the immune system and decrease the inflammation caused by the LPS challenge. Previous study by Soliman et al and Chen et al have demonstrated that acetate supplementation is able to increase the production of IL-4 and IL-10, diminishing the inflammatory processes in a neuroinflammation model and in neutrophils (Soliman, Puig et al. 2012, Chen, Wu et al. 2016). Thus, our current study corroborates this hypothesis of the protective effects of ferulic acid and other metabolites modulating the host metabolism and immune system. Furthermore, we were able to demonstrate that the growth of this bacterium alone *in vitro* is able to achieve similar results in human tissue culture.

Although human tissue culture provides invaluable answers regarding the effect of metabolites, the complexity of an *in vivo* system is still unattainable by these methods. To determine if the *in vitro* culture and ferulic acid accumulation by *Bacteroides* is able to reach significant levels to modulate the immune system in a conventional mice model, we gavaged mice with different treatments. In fact, the treatments containing high concentrations of ferulic acid (positive control, wheat bran and destarched wheat bran spent medium) were able to increase cell differentiation, increasing macrophage and granulocytes cells, and important TH1-related cytokines IL-2 and IL-12 both in spleen and mesenteric lymph nodes. The results suggest that the ferulic acid may be absorbed by the intestinal epithelium and reach the bloodstream and elicit an immunological response in the host. Although granulocytes and macrophages are commonly associated with the humoral immune system, these cells are responsible for presenting antigens to T cells

and training the adaptive immune system (Sallusto and Lanzavecchia 1994). Furthermore, increased expression of IL-2 has been associated with differentiation of naïve T cells into regulatory T cells (Treg) (Lee, Wang et al. 2015). Treg cells are responsible for modulating the adaptive immune system, by modulating the activity of T effector cells and diminishing inflammatory response. The increased expression of IL-12 is key for the differentiation of T naïve cells to T helper cells 1 (Th1) (Manetti, Parronchi et al. 1993), increasing the adaptive immune response towards intracellular pathogens and restoring the balance between Th1/Th2 cells. These results show that the culture of *Bacteroides* in complex arabinoxylan substrates is able to increase the ferulic acid pool in the gastrointestinal tract to be absorbed by the epithelial lining, modulating the adaptive immune response.

In summary, our work demonstrates that *Bacteroides* species are able to degrade esterified arabinoxylan coming from a rich whole cereal grain diet, while generating short-chain fatty acids that are known to modulate energy and host balance (Clausen and Mortensen 1995, Topping and Clifton 2001, Le Poul, Loison et al. 2003). Furthermore, the activity of the esterase-enriched cluster employed by *Bacteroides* species is able to release this important phenolic compound, ferulic acid, and increasing the availability to the host and microbial community. This phenolic compound can be absorbed by the host intestinal epithelium and modulate the adaptive immune system (Figure 2.17) (Ohnishi, Matuo et al. 2004, Jin Son, C et al. 2010, Lee, Wang et al. 2015). A recent study has also demonstrated the ability of phenolic compounds towards decreasing glycemic index in diabetes mouse models, as well as allergies (Lee, Wang et al. 2015). Thus, the activity of these bacteria can have a larger impact than anticipated by this study. Furthermore,

previous studies have demonstrated that the gut microbiome is able to metabolize ferulic acid into phenolic derivatives through the phenylpropionic acid pathway (Russell, Duncan et al. 2013). Thus, more need to be understood regarding the impact of these derivatives towards the host metabolism.

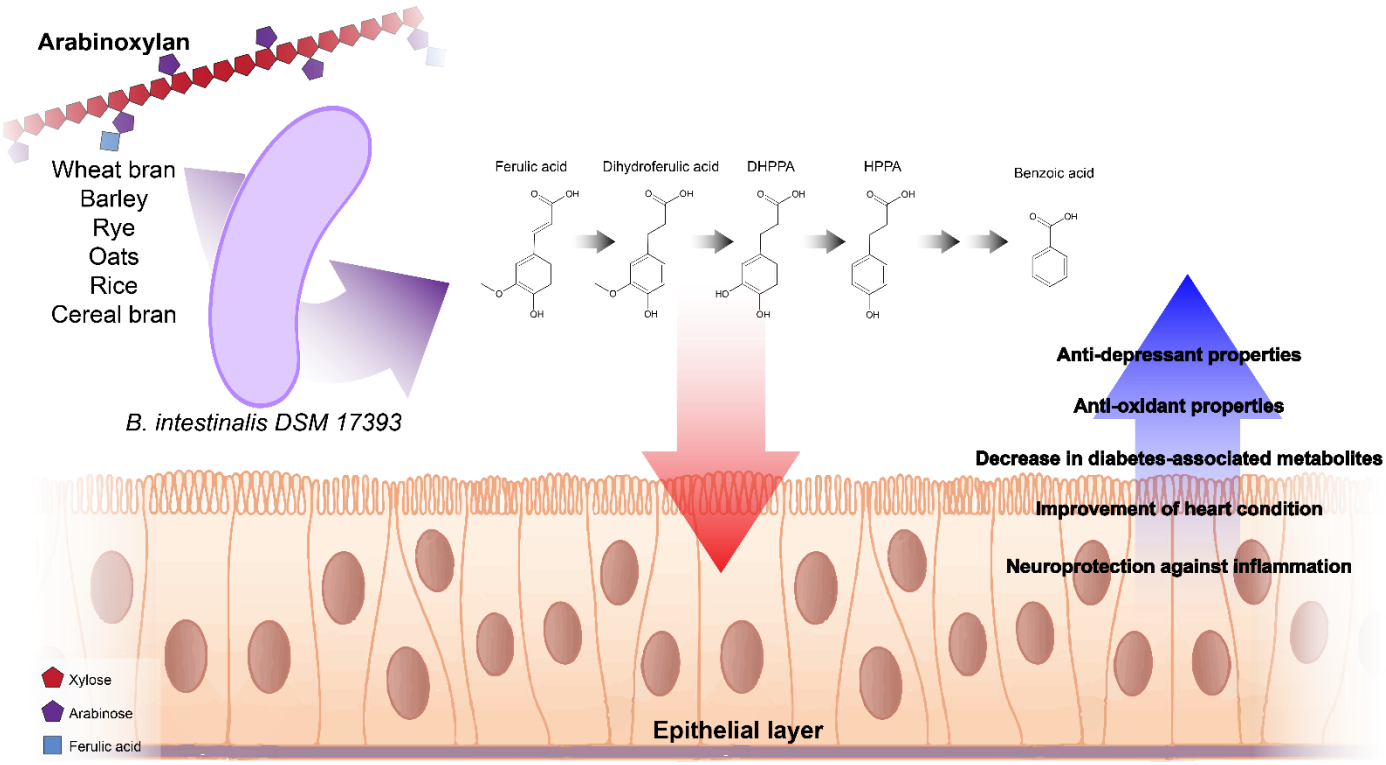


Figure 2.17 - **Fate of ferulic acid released from an arabinoxylan-rich diet with esterase-enriched cluster activity.** Arabinoxylan coming from a whole cereal diet can be depolymerized by *Bacteroides* species encoding the esterase-enriched cluster, increasing the pool of ferulic acid available in the gastrointestinal environment. Ferulic acid can be absorbed by the host or metabolized into ferulic acid derivatives by the microbiome, leading to modulation of the host homeostasis.

2.4 Materials and methods

Bacterial strains and materials. *Bacteroides intestinalis* DSM17393 [23] was obtained from the Deutsche Sammlung von Mikrooganismen und Zellkulturen or German Collection of Microorganisms and cell Cultures (DSMZ, Braunschweig, Germany). The NEB® 5-alpha Competent *Escherichia coli* (High Efficiency) and Phusion® High-Fidelity DNA Polymerase were purchased from New England Biolabs (Ipswich, MA). The *E. coli* BL21-CodonPlus(DE3) RIL competent cells were obtained from Agilent (Santa Clara, CA). The pET-28a plasmid was purchased from Novagen (Merck, Darmstadt, Germany) with the kanamycin resistance gene replaced by that conferring ampicillin resistance. The QIAprep Spin Miniprep kit and QIAquick PCR Purification Kit were obtained from Qiagen, Inc. (Valencia, CA). Amicon Ultra-15 centrifugal filter units with 10 kDa molecular mass cutoffs (MMCOs) were obtained from Millipore (Billerica, MA). Soluble wheat arabinoxylan with medium viscosity (WAX), insoluble wheat arabinoxylan (Ins. WAX) and xylo-oligosaccharides were obtained from Megazyme (Bray, Ireland), and oat spelt xylan (OSX) was purchased from Sigma-Aldrich (St. Louis, MO).

Growth conditions. The growth conditions were conducted inside an anaerobic chamber with an 85% N_2, 10% CO_2, and 5% H_2. *Bacteroides intestinalis* were cultured from a single colony into supplemented Brain-heart infusion medium (BHIS). For all growth experiments, 100 µL of cells were inoculated into defined medium containing each different carbon source and passed three times to remove any residues from the growth in BHIS medium.

Gene cloning, expression, and protein purification. The Polysaccharide-Utilization Loci DataBase (PULDB) server (http://www.cazy.org/PULDB/) was used to

identify the predicted esterase-enriched PUL in the *B. intestinalis* genome. Signal peptides and lipoprotein signal sequences were predicted using SignalP v4.1 (http://www.cbs.dtu.dk/services/SignalP/) and LipoP v1.0 (http://www.cbs.dtu.dk/services/LipoP/), respectively. Each putative esterase (BACINT_01033, BACINT_01038, BACINT_01039, and BACINT_01040) and putative glycoside hydrolase (BACINT_01035, BACINT_01041, BACINT_01042, and BACINT_01043) were amplified from *B. intestinalis* DSM 17393 genomic DNA via PCR using the NEB Phusion® High-Fidelity DNA Polymerase. The putative bifunctional enzyme encoded by HMPREF1016_RS0111555 was amplified from *Bacteroides* egghertii from genomic DNA by PCR. The PCR primers used for gene amplification are listed in the Supplementary Table 1. The PCR products were purified and digested with NdeI and XhoI. The PCR products were cloned in the modified pET-28a expression vector at the NdeI/XhoI in frame with the hexahistidine encoding sequence. The ligated products were transformed into NEB® 5-alpha Competent *E. coli*, and the transformants were plated in lysogeny broth (LB) agar plates supplemented with 100 µg/mL ampicillin. The recombinant plasmids harboring the correct inserts were confirmed by DNA sequencing (W. M. Keck Center for Comparative and Functional Genomics, University of Illinois). For gene expression, the recombinant plasmids containing individual genes were introduced into *E. coli* BL21-CodonPlus(DE3) RIL competent cells by heat shock transformation and cultured overnight on LB agar plates supplemented with ampicillin (100 µg/ml) and chloramphenicol (50 µg/ml) at 37 °C. After 14 hours, individual colonies containing different genes were picked and inoculated into 10 ml fresh LB medium containing the same concentrations of ampicillin and chloramphenicol and incubated at 37 °C with

Table 2.1 – Primers used for cloning *B. intestinalis* genes into pET28a (+)

Primer (5'-3')	Gene	Sequence[a]
Bi1033For	BACINT_01033	catatgCAACAGCAAGATTTTCCGGCAGGAAC
Bi1033Rev		ctcgagTTATTTCGTTTTAAATAATAGGGGAGCAAATTCTTTCAGGC
Bi1035For	BACINT_01035	catatgCAAATCGGCACTCCATACATCCACGATC
Bi1035Rev		ctcgagCTAATGGTCGCGGAAATTCCATTTGGAATTG
Bi1038For	BACINT_01038	catatgTTGAATAGAATGAAAAAGCTGTTATTATTTATCGCATGCTTG
Bi1038Rev		ctcgagTTACTTCTTAAATAAATTCGGTAAAAATTCATTCAGACATCTGC
Bi1039For	BACINT_01039	catatgCAGACAGTGGAGGATTTCAAACCATCG
Bi1039Rev		ctcgagTTATTTAAAAAGAAGCGGAGCAAACTCATTCAATGC
Bi1040For	BACINT_01040	catatgCAGATTACGCAATGGACTGATATCAACTATGC
Bi1040Rev		ctcgagTTAATAAAGGGTATAAAACTGTTTTTGAGGAGGATTTTTC
Bi1041For	BACINT_01041	catATGAAGATACTGTTTCATTTCACAATAACTCTGTTCG
Bi1041Rev		ctcgagCTAAGGATGATATTTCCCCAGTATTCTAATTTCGTCTC
Bi1042For	BACINT_01042	catatgCAAACATTGCCGTATCAGAATCCTGAACTAAG
Bi1042Rev		ctcgagTTATTGTAAAGTGACTTTGACAGATTGCAGGTC
Bi1043For	BACINT_01043	catatgCAGAATCCCATTATTACGGATCAGTTCACTG
Bi1043Rev		ctcgagTTATTGAAAACTGATCCAGTCGATTTCAACTTTACC
BeGH43/FaeFor	HMPREF1016_RS0111555	ccgggatccCAAAAGCCTGCAACTAATCCTGTGA
BeGH43/FaeRev		ccgctcgagTCATTTAATGTCATCGCATTTTATCGGCC

[a] Lower case sequence demonstrate the included restriction enzyme sites.

vigorous shaking (250 rpm/min) for 6 hours. Each 10 ml culture was then transferred into a 1 liter LB medium in a 2.8 liter Fernbach flask and culturing continued at 37 °C until the optical density at 600 nm (OD_{600}) reached 0.6. The gene expression inducer isopropyl β-D-thio-galactopyranoside (IPTG) was added to the culture at a final concentration of 0.1 mM to induce recombinant gene expression. The culturing was continued for 16 hours at a temperature of 16 °C with shaking at a speed of 200 rpm/min. The cells were harvested by centrifugation and resuspended in ice-cold binding buffer (pH 7.5, 50 mM Tris-HCl, 300 mM NaCl) and ruptured by passage through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Ontario, Canada). Each clarified supernatant, containing the cell contents, was obtained after centrifugation at 12,857 x *g* for 30 mins at 4 °C and loaded into an ÄKTAxpress fast protein liquid chromatography (FPLC) system equipped with a HisTrap™ 5 mL (GE Healthcare, Piscataway, NJ). The proteins bound to the column were washed with five column volumes of binding buffer and two columns volumes of binding buffer supplemented with 10 mM imidazole. The proteins were eluted in a gradient of binding buffer and elution buffer (pH 7.5, 50 mM Tris-HCL, 300 mM Sodium Chloride, 500 mM imidazole) and collected in 1.5 mL fractions. The purest fractions were polled, concentrated into a final 2 mL volume and centrifuged at 25,000 x *g* for 5 mins to precipitate any denatured proteins. The recombinant protein were further purified through size exclusion chromatography. The proteins were loaded in ÄKTAxpress fast protein liquid chromatography (FPLC) system equipped with a HiLoad 16/60 Superdex 200 column (GE Healthcare, Piscataway, NJ). The chromatography were solved using a isocratic flow of a buffer composed of 50 mM Tris-HCL, 150 mM Sodium Chloride, pH 7.5. Samples from eluted fractions were analyzed by 12% sodium dodecyl sulfate-

polyacrylamide gel electrophoresis (SDS-PAGE) according to the method of Laemmli[38], and protein bands were visualized by staining with Coomassie brilliant blue G-250. The highly purified protein fractions were used for downstream applications.

Substrate specificity of glycoside hydrolases in different xylan substrates. To investigate the enzymatic specificity of the four glycoside hydrolases on polysaccharide substrates (wheat arabinoxylan, rye arabinoxylan, and glucuronoxylan) the substrate specificity for each enzyme was determined individually at pH 6.5 and 37°C, with a final enzyme concentration of 500 nM. Fifty microliter aliquots were removed at end point assay (16 hours) from the reaction mixtures, and the concentrations of reducing ends released with time were assayed using the *para*-hydroxybenzoic acid hydrazide assay method with known xylose concentrations used in plotting a standard curve. For BACINT_01042 and BACINT_01043, a similar end point assay was performed against xylo-oligosaccharides and arabino-oligosaccharides. The end products were analyzed by HPAEC-PAD on an ICS-5000 system (Thermo Scientific Dionex, Sunnyvale, CA) equipped with a CarboPac PA-100 column (250 mm × 2 mm; Thermo Scientific Dionex). The flow rate was 0.25 mL/min, and a gradient composed of the following eluents were used at 25 °C: (A) 0.1 M sodium hydroxide and (B) 0.1 M sodium hydroxide + 0.1 M sodium acetate.

Carbohydrate esterases activity towards complex substrates. The release of ferulic acid from the complex substrates, insoluble wheat arabinoxylan and sugar beet pulp, were monitored by incubating different concentrations (0.5, 1.0, and 1.5 µM) of each putative carbohydrate esterase with 10mg/mL of each substrate (in 50 mM sodium citrate, 150 mM NaCl pH 6.5, pre-swollen for 10 minutes at 80°C) for 2 hours at 37°C and 1200

rpm. At the end of the reaction, each sample was heated at 99°C for 10 minutes, cooled to room temperature, one volume of methanol was added and vortexed for 30s. The reaction suspension was centrifuged at 15.000 rpm for one minute, and the clear supernatant analyzed by HPLC-DAD system (LC-20 AD pumps, SIL-20A autosampler, SPD-M20A PDA detector; Shimadzu, Kyoto, Japan) equipped with a Luna C18 column (250 mm × 4.6 mm i.d., 5 µm particle size; Phenomenex, Torrance, CA). The flow rate was 0.5 mL/min, and the following gradient composed of (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid was used at 45 °C: 0–20 min, isocratic 85% A and 15% B; 20–35 min, linear to 100% B; 35–40 min, isocratic 100% B; and 40–45 min, isocratic 85% A and 15% B. UV detection was carried out at 325 nm, and a standard curve was used to determine the ferulic acid content of the hydrolysates.

FA-FAX-FAXX isolation. The preparation of de-starched wheat bran and purification of feruloylated oligosaccharides it was done following a previous protocol (Daniel). In summary, 60g of milled wheat bran (<0.5 mm) was suspended in 600mL of phosphate buffer (80 mM, pH 6.2) and incubated with 4.5 mL of thermostable α-amylase (Termamyl 120 L, EC 3.2.1.1, from *Bacillus licheniformis*, 20,000–60,000 U/mL) at 92°C for 20 minutes. The insoluble fraction was obtained by centrifugation at 4.000 RPM for 10 minutes; washed twice with water, twice with ethanol, and acetone twice; and dried at 40°C. For the isolation of feruloylated oligosaccharides from the destarched wheat bran, 30g of the destarched wheat bran were suspended in 3L of water and autoclaved for 40 minutes. Subsequently, 1.5g of Driselase enzyme preparation was added to the suspension and incubated at 37°C with 80 rpm for 48h. The mixture was then subjected to 95°C for 10 minutes and any undigested product removed by centrifugation. The non-

phenolic products were removed by passing through an Amberlite XAD-2-column (bed volume: 55 x 2.6 cm) rinsed with methanol and preconditioned with water. The hydrolysate was applied to the column, which was then washed with 500mL of water. The feruloylated oligosaccharides were eluted using 500mL of methanol/water (50:50, v/v). Subsequently, the methanol was removed by evaporation, and the aqueous eluate was freeze-dried.

Preparation of acetylated xylan. The acetylated xylan was prepared in accordance with previous reports (Johnson, Fontana et al. 1988). In summary, oatspelt xylan was slowly added to 250 mL of dimethyl sulfoxide at gentle stirring at room temperature until 4% wt/volume. After the xylan is evenly suspended, the preparation was heated to 55°C over a 20 minutes period until solubilized. After solubilized, 200 mL of acetic anhydride pre heated to 60°C, was added with stirring over five minutes. The reaction mixture was placed in a dialysis tubing and dialyzed for 5 days under running water at 4°C. Following, 24hours of dialysis against ddH2O. After dialysis, the substrate was freeze dried.

Esterase substrate specificity and specific activity towards ester-linked substrates. The substrate specificity of every esterase was assess by incubating 200 µM for the purified feruloylated arabinoxylan oligosaccharides FA, FAX, ,FAXX, FAA, FAAA, FGG, and FAXG with 50 nM of each putative carbohydrate esterase (BACINT_01033, BACINT_01038, BACINT_01039, and BACINT_01040) in citrate buffer ((50 mM sodium citrate, 150 mM NaCl, pH 6.5) at 37°C for 10 minutes. The enzymes were inactivated by heating the reaction at 95°C for 10 min, and the solutions were analyzed by HPLC-DAD under the same conditions as described above. The feruloylated oligosaccharides and

the hydrolysis products were detected at 325 nm. For the acetyl esterase activity, the substrates (0.5% acetylated oat spelt xylan, 0.4 mM tetra-O-acetyl-xylopyranose) were incubated with each putative carbohydrate esterase in citrate buffer pH 6.5 at 37°C for 1 hour, and the released acetic acid was measured using an acetic acid detection kit (Megazyme, Bray, Ireland) according to the manufacturer's instructions. The reduction of NADH was monitored continuously at an absorbance of 340 nm using a Synergy 2 microplate reader (BioTek, Winooski, VT) using the path length correction feature. For the hydrolysis of the artificial substrates, Methyl-ferulate and *p*NP-acetate, 50 nM of enzyme was incubated with 1 mM of each substrate, in 0.1M MOPS buffer pH 7.5 at 37°C. The reaction was continually monitored at 350 nM for methyl-ferulate and 400 nM for *p*NP-acetate, using methyl-ferulate and *p*NP, respectively, for the standard curve.

Optimal pH and temperature esterases. The optimal pH of the esterases Bi1033, Bi1038, Bi1039, and Bi1040 were determined by incubating each enzyme at 50 nM final concentration with 2 mM *p*NP-acetate at 37°C in a range of pH from 4-6.5 (50 mM sodium citrate and 150 mM NaCl) and 6-8.5 (50 mM sodium phosphate and 150 mM NaCl) , and the released *p*NP was continuously monitored by determining the absorbance at 400 nm using the Cary 300 UV-Vis spectrophotometer (Agilent, Santa Clara, CA). The optimal temperature of each enzyme was measured in its corresponding optimal buffer at temperatures ranging from 20 to 75°C.

Synergistic activity of glycoside hydrolases and carbohydrate esterases. To analyze the synergistic effect of glycoside hydrolases (homologous or heterologous) and the carbohydrates esterases, the arabinoxylan substrates (Insoluble wheat arabinoxylan and destarched wheat bran, each enzyme was individually or combinatory incubated with

0.5% wt/vol final concentration of the substrates at pH 6.5, 37°C for 14 hours. The ferulic acid and reducing sugar ends released were measured as described above.

Transcriptional analysis. Cells grown in the monosaccharides xylose:arabinose and insoluble wheat arabinoxylan were collected in two volumes of RNAprotect Bacteria Reagent (Qiagen, Germantown MD). Cells were pelleted by centrifugation at 10,000 x g for 10 minutes and resuspended in lysis buffer (10 mM Tris, 1 mM EDTA, 1 mg/ml lysozyme, 0.1 mg/ml Proteinase K) for 10 minutes. Total microbial RNA was extracted using the RNeasy Mini Kit from Qiagen following the manufacturer's protocol with an on column DNase treatment step. RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara CA) with all samples achieving a RNA Integrity Number (RIN) of 9 or higher. The High-Throughput Sequencing and Genotyping Unit of the Roy J. Carver Biotechnology Center at the University of Illinois Urbana Champaign performed high throughput sequencing using an Illumina 4000 sequencing platform.

Resulting sequence data was analyzed for differentially expressed genes following a previously published protocol (Anders, McCarthy et al. 2013). Briefly, reads were filtered for quality using Trimmomatic (Bolger, Lohse et al. 2014). Reads were aligned to the genome using BowTie2 (Langmead and Salzberg 2012). Reads mapping to gene features were counted using htseq-count (Anders, Pyl et al. 2015). Differential expression analysis was performed using the edgeR package in R and the TMM method was used for library normalization (Robinson, McCarthy et al. 2010). Coverage data was visualized using Integrated Genome Viewer (IGV) (Thorvaldsdottir, Robinson et al. 2013).

RNA Extraction and reverse transcription quantitative PCR (RT-qPCR). The cells were grown as described above. For growth condition, 5mL of cells at mid-log phase

were rapidly mixed with RNAprotect bacterial reagent as described by the manufacturer and was used for subsequent RNA isolation. RNA was isolated using the RNeasy mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Contaminating DNA was removed in-column using NEB DNAse I (Ipswich, MA) and quality of the RNA was assessed by using a Bioanalyzer 2100 with a RNA 6000 Nano Assay Reagent kit from Agilent (Santa Clara, CA). RNA quantity was measured using the nanodrop (Agilent) and Qubit (Thermo Fisher Scientific). RT-qPCR was performed using a Thermo light cycler 480 (Roche, San Fransisco CA) and SYBR green master mix (Roche, San Fransisco CA). cDNA was generated from 1 µg total RNA using a transcriptor High fidelity cDNA synthesis kit (Roche, San Fransisco CA) according to the manufacturers manual. Synthesized cDNA was 100 x diluted for qRT-PCR (corresponding to a start amount of 10 ng RNA).The relative expression of SusC/SusD pair in the esterase-enriched clusters of *Bacteroides intestinalis*, *Bacteroides cellulosilyticus* and *Bacteroides oleiciplenus* in the xylan-related substrates (Destarched wheat bran, soluble wheat arabinoxylan, insoluble wheat arabinoxylan, and feruloylated oligosaccharides) culture conditions was compared to a xylose:arabinose mixture at a mid-log growth. The designed specific primers are listed at Supplementary Table 2. The fold change in gene expression (xylan substrates *versus* monosaccharides mixture) was calculated from three biological replicates (+ three technical replicates) using the 16S rRNA reference gene for normalization. Log2 fold-change at mid-log phase were considered as significantly different at $p < 0.01$ (Student's t-test).

Comparative genomic analysis. To understand the prevalence of an esterase-enriched cluster similar to *B. intestinalis* operon, a comparative genomics approach was

Table 2.2 – Primers used for qPCR in *B. intestinalis*, *B. cellulosilyticus*, and *B. oleiciplenus*

Primer (5'-3')	Gene	Sequence
Bi1036For	BACINT_01036	TGATTCTAACTACACTCTCTTTGGTG
Bi1036Rev		TTGACAGGAGTAACGTTCATGTAAGC
Bi1037For	BACINT_01037	TATCAATCATGAGTTTGCTCACAAAATG
Bi1037Rev		TTCGTCAAATTCTTTATCTGCGCG
Bi16sFor	Bi16s rDNA	GGAGCGTAGGCGGATTATTAAG
Bi16sRev		GGAGCGTAGGCGGATTATTAAG
Bc2149For	BACCELL_02149	GAATGGGCAGGTAACAGAAA
Bc2149Rev		GGATGGCAACCGTACAGATAG
Bc2148For	BACCELL_02148	CACTCCAGACGAGTTCTTATAC
Bc2148Rev		GGCCTTCTGTACTCTTGATACC
Bc16sFor	Bc16s rDNA	AGCAACACAATGCTATG
Bc16sRev		CACGTAAACCACTTTCTT
Bo2534For	HMPREF9447_02534	CGTATCCGGCTTCACCTATTC
Bo2534Rev		GTCGAAACCTACACCCATATCC
Bo2533For	HMPREF9447_02533	GATGCGCTGGCATAACAATAC
Bo2533Rev		CTCTACTCGGAATCGGAAGAATG
Bo16sFor	Bo16s rDNA	CCATTCATTGGGCATAG
Bo16sRev		CGTACTTTCTTACCGATAC

determined. The usage of >20 genomes of *Bacteroides* common in the human gut were employed to identify the presence of an esterase-enriched cluster using Geneious version 9.0.2 for further analysis. The genome files were aligned using Mauve version 2.3.1 plugin in Geneious software to perform a genome-wide alignment. This alignment used *B. intestinalis* as a reference genome in order to identify the potential esterase-enriched clusters.

Growth analysis in complex insoluble substrates, wheat bran and destarched wheat bran. To determine the ability of *B. intestinalis* to grow on these complex substrates, we determine the initial and final protein concentration in the medium by protein quantification and fermentation products analysis. The total protein quantification was performed using a Bradford microtiter assay (Biorad) according to the manufacturer's instructions. For the fermentation products analysis, in the initial and final supernatant acetate, succinate, and proprionate concentrations were determined by high-performance liquid chromatography (HPLC, Agilent Technologies 1200 Series, Mississauga, CA) equipped with a refractive index detector using a Rezex ROA-Organic Acid H^+ (8%) column (Phenomenex Inc., Torrance, CA). The column was eluted with 0.005 N of H_2SO_4 at a flow rate of 0.6 ml/min at 50°C.

Crystallography studies. The crystallization conditions for Bi1033, Bi1039, and BeGH43/FAE were evaluated by using the sparse matrix sampling method and commercial screens. For preparation of seleno-methionine-labeled Bi1033 and Bi1039, transformed *E. coli* BL21 (DE3) cells were grown in M9 medium (containing 1 mM $MgSO_4$, 0.4% glucose, 0.5 mg/L thiamine, and 100 µg/mL ampicillin) at 37 °C to an OD_{600} of 0.4. Following the addition of an amino acid mixture (lysine, threonine, phenylalanine, leucine,

isoleucine, and valine) and seleno-methionine, protein overexpression was induced by the addition of IPTG. The cultures were then grown for 24 h at 16 °C, followed by purification of the protein as described above. Crystals of the seleno-methionine-labeled Bi1033 and Bi1039 were prepared by using the hanging drop diffusion technique. The protein [2 µL, 20 mg/mL, 50 mM Tris,150 mM NaCl (pH 7.5)] was mixed with 2 µL of precipitant solution from Hampton Research Crystal screen condition 41 [70 mM Ammonium sulfate, 80 mM Sodium Cacodylate pH 6.5, and 16% PEG4000].

Phylogenetic analysis of Carbohydrate Esterases. All carbohydrate esterases associated with Polysaccharide Utilization Loci (PUL) used in this study were extracted from the PUL database on 20 February 2018 (~500 sequences). The CE domains from the proteins were used for phylogenetic tree, bifunctional esterases were separated based on Conserved Domain Database and used as single domains. Duplicated sequences were removed and sequence alignment was performed by MUSCLE (Edgar 2004). To generate high-quality and relevant alignments, MAFFT was used to iteratively remove highly dissimilar sequences (Katoh, Misawa et al. 2002). Total of ~500 sequences remained which then was used to generate a phylogenetic tree using FASTTree based on the midpoint root method (Price, Dehal et al. 2010). The tree and subtrees were visualized and edited with iTOL v3 (Letunic and Bork 2016). The annotations of the clusters are based on the family associated to enzymes.

Immunological studies of human tissue culture. Caco-2 cells were cultured in a 12-well plate in DMEM +20% FBS. The HIEC-6 cells were cultured until semi-confluent in DMEM +5% FBS supplemented with 10 ng/mL epidermal growth factor (EGF) and 0.2 IU/mL insulin. The cells were serum starved overnight and half a volume of condition

(wheat bran blank, wheat bran spent medium, destarched wheat bran blank, destarched wheat bran spend medium, and trans-ferulic acid to the final concentration of 0, 1, 10, and 100 μM. The cells were incubated during an hour with each condition and challenged to oxidative stress with 10 μg/mL E. coli lipopolysaccharide (LPS). The cells were cultured for 24 hours and the supernatant was collected for ELISA and cells for RNA extraction and RT-qPCR.

CHAPTER 3: UTILIZING A HYBRID TWO-COMPONENT SYSTEM SENSOR DOMAIN SIGNATURE TO PREDICT CARBOHYDRATE UTILIZATION IN BACTEROIDETES

3.1 Introduction

The human gut microbiome is a complex, diverse, yet stable environment composed of trillions of bacteria. Despite its diversity, members of the Bacteroidetes and Firmicutes phyla dominate the colon environment (Turnbaugh, Backhed et al. 2008, Turnbaugh, Hamady et al. 2009). One of the main determinants for bacteria to thrive in the human gastrointestinal tract is the ability to access nutrients that evade the host digestive system. Dietary fibers are polysaccharides, commonly from plant cell wall, that the host enzymatic machinery is unable to degrade, reaching the colon largely intact and prone to microbial degradation (Martens, Lowe et al. 2011). The Bacteroidetes are equipped with a unique enzymatic machinery that degrades, transports, and senses the dietary and host glycan polysaccharides to carry out fermentation generating energy and several metabolites that influence both nutrition and host health (Martens, Chiang et al. 2008). The Bacteroidetes encode these polypeptides in large clusters on their genomes, known as polysaccharide utilization loci (PUL). In fact, currently there are around 5,000 different PULs deposited at the Polysaccharide Utilization Loci Database (PULDB) with similar organization of carbohydrate sensing and transport, but widely varied polysaccharide targets. Despite the large repository of PUL at this database, the activity of the majority of PULs are still unknown.

The dominant mechanism of sensing by prokaryotes are a two-component system composed of a membrane-bound sensor histidine kinase and a cytoplasmic response regulator. This mechanism allows sensing of extracellular stimuli coupled to adaptive

responses by the bacteria (Hoch and Silhavy 1995). Interestingly, Bacteroidetes seem to have evolved a similar system, hybrid two-component system, where one gene encodes a polypeptide composed of the periplasmic sensor domain combined with the cytoplasmic response regulator (Lowe, Baslé et al. 2012). Corresponding to the classical two-component system, the HTCS sensing carbohydrate generates a conformational change that allows a phosphorylation cascade that leads to activation of gene transcription (Schwalm, Townsend et al. 2017). Furthermore, the sensing mechanism of HTSCs require the action of PUL encoded genes to cleave and transport the extracellular polysaccharide to the periplasmic region. Thus, deciphering the mechanism and specificity of HTCS sensor domain will shed light onto how a PUL targets polysaccharide.

In this work, we aim to develop a rapid way to identify and assign function to unknown PULs based on bioinformatics and biochemical analysis. Based on bioinformatics analysis, we predict that most HTCS are composed of a signal peptide, a carbohydrate sensor domain, a Y_Y_Y module, a transmembrane region, a histidine kinase module, a histidine ATPase module, a response regulator module, and a helix-turn-helix domain. Seemingly the cytoplasmic region, especially the modules related to the signal transduction, are conserved between the different HTCS. This leads to the hypothesis that the sensor domain possesses the most variability to sense different sugars and activate gene transcription. Furthermore, *susC/susD*-like genes seems to be core genes for the structure of PULs, analysis of several different PULs has demonstrated the presence of a potential endo-acting enzyme near the *susC/susD* pair. This phenomenon leads to the hypothesis that this protein may act as a scouting enzyme. In this study, we biochemically characterize the activity of two scouting enzymes,

demonstrating that its end-products released by the reaction with cognate substrates are sensed by its corresponding HTCS sensor domain. Therefore, potentially allowing to determine the target PUL polysaccharide specificity. To determine if our working hypothesis allow us to understand arabinoxylan and arabinan degradation by *Bacteroides intestinalis*, we performed RNAseq in this bacterium grown on the cognate monomers and polysaccharides to compare transcriptomic responses. As expected, this bacterium was able to sense the oligosaccharide generated by the scouting enzyme of each PUL, and activate gene transcription.

To further understand the action mechanism of the arabinoxylan HTCS, we solved the structure of the sensor domain region and by site-directed mutagenesis identified the amino acids responsible for oligosaccharide binding. The results of the present work provides a way to rapidly identify and assign function to different polysaccharide utilization loci.

3.2 Results

Scouting enzyme and hybrid two-component system sensor domain substrate specificity allows rapid identification of polysaccharide utilization loci targets. The polysaccharide utilization loci database houses roughly 5,000 different hybrid two-component systems deposited, with the majority of PULs with no known target polysaccharide assigned. Recent transcriptomic analysis of *Bacteroides* species has demonstrated that *susC/susD*-like genes are highly expressed in the presence of the polysaccharide that the PUL targets (Martens, Chiang et al. 2008, Zhang, Chekan et al. 2014). Furthermore, the enzyme next to the *susC/D* pair seems to follow a similar expression level. This led to the hypothesis that this enzyme acts as a scouting enzyme,

that initially detect and degrade the cognate polysaccharide into oligosaccharides to be transported into the periplasmic region, allowing the HTCS sensor domain to sense and increase transcription of the entire PUL. To assess this hypothesis, we identified a polysaccharide utilization loci that seems to target arabinoxylan substrates (Figure 3.1A). We cloned, heterologously expressed and purified the putative endoxylanase encoded by BACINT_RS19655 and the sensor domain region of the HTCS close to homogeneity (Figure 3.1B). The product encoded by the scouting gene demonstrated endoxylanase activity towards wheat arabinoxylan. HPLC-HPAED analysis of the reaction products released by incubation of the arabinoxylan-scouting enzyme demonstrated a wide range of xylooligosaccharides from xylose to xylopentaose (Figure 3.1C). This result, combined with the presence of signal peptide allows us to hypothesize that the scouting enzyme is on the extracellular region, and that it is able to cleave the polysaccharide in close proximity to the bacterium. To assess the ability of the HTCS sensor domain to bind to the end products of the scouting enzyme reaction, we used isothermal titration calorimetry to measure the binding of different xylo-oligosaccharides to the sensor domain. The sensor domain exhibited binding to oligosaccharides from xylotetraose to xylohexaose, with the largest affinity towards the larger xylopentaose and xylohexaose (Figure 3.1D). Despite the fact that xylohexaose was not detected as an end-product of the scouting enzyme during incubation with the polysaccharide, this reaction would initially generate larger fragments that can be rapidly transported to the periplasmic region and sensed by the HTCS. This can lead to up-regulation of the cluster during growth in arabinoxylan polysaccharide by the bacteria. In a similar way, a PUL predicted to target arabinan possesses a scouting enzyme as well as a hybrid two-component system regulator

(Figure 3.2A). Therefore, in order to understand if the concept of scouting enzyme is wide and prevalent in different PULs, the putative endo-arabinanase and the sensor domain of the regulatory protein were expressed and purified close to homogeneity, as shown in the SDS-PAGE analysis (Fig 3.2B). The product encoded by this gene demonstrated endo-1,5-α-arabinofuranosidase activity towards both branched arabinan and debranched arabinan substrates. HPLC-HPAED analysis of the products released by the incubation of BACINT_02777 and branched arabinan demonstrated a wide range of arabino-oligosaccharides from arabinobiose to arabinopentaose (Fig 3.2C). The presence of a signal peptide with a wide variety of end products released by BACINT_02777, led to the hypothesis that this enzyme breaks down arabinan into products that are recognized by the HTCS. To test this hypothesis, isothermal titration calorimetry (ITC) was used to determine the binding affinity of the end-products by the sensor domain. The results showed a lower affinity for arabinopentaose, while a higher affinity was detected for arabinohexaose (Fig 3.2D). Although arabinohexaose was not detected in our analysis, the breakdown of arabinan would initially generate larger fragments that can be detected by the sensor domain to lead to the up-regulation of this PUL.

Transcriptomic analysis of *Bacteroides intestinalis* under different substrates supports scouting enzyme concept. To test the utility of the concept of using the HTCS sensor domain and the scouting enzyme as a prediction tool for carbohydrate metabolism by the Bacteroidetes, we compared growth of *Bacteroides intestinalis* on the two polysaccharides, arabinoxylan and arabinan, with growth on their

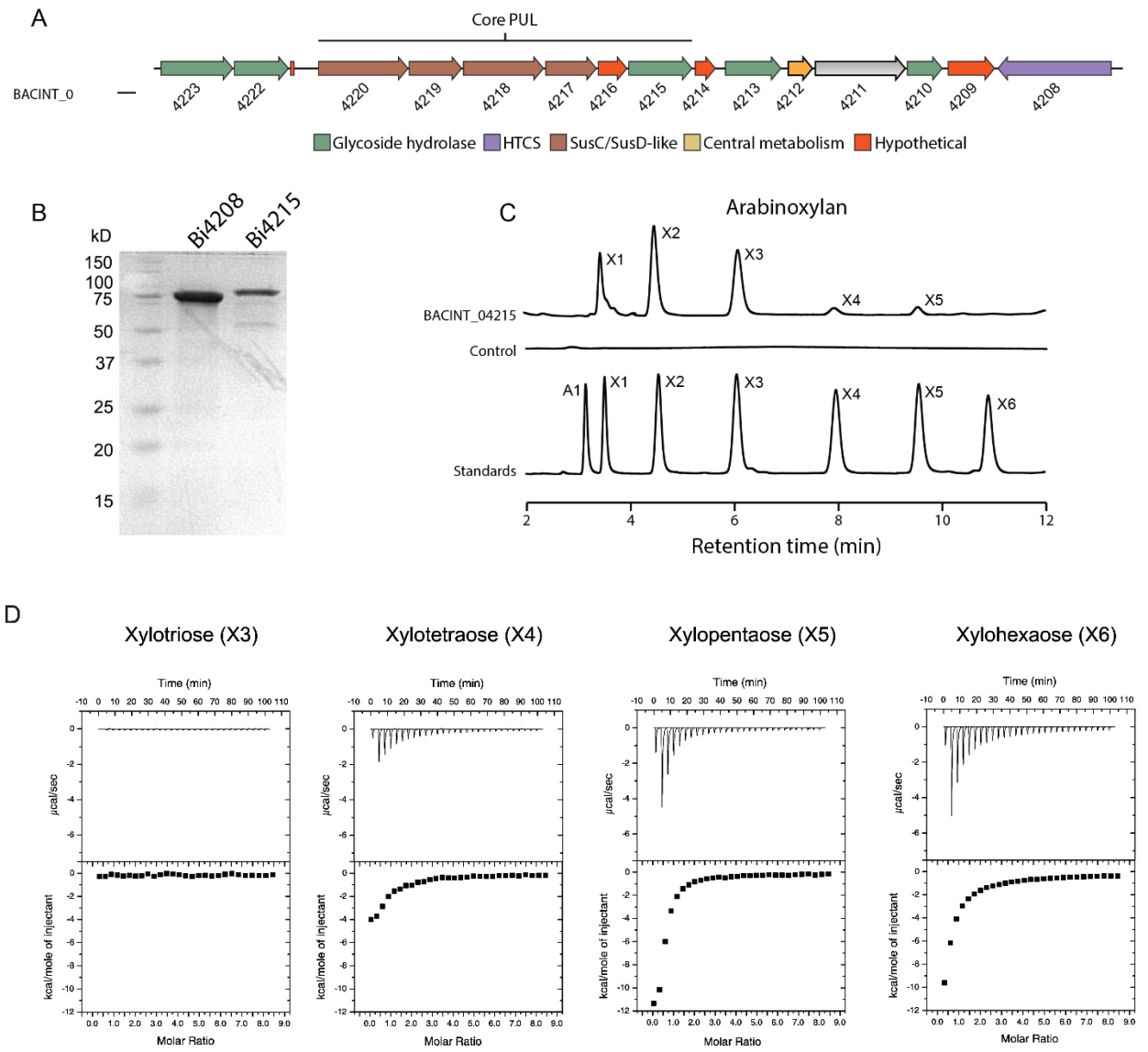


Figure 3.1 - **Identification of an arabinoxylan polysaccharide utilization loci (PUL) in *Bacteroides intestinalis*.** (A) Polysaccharide utilization loci map of the arabinoxylan. PUL is color-coded based on predicted functions: Glycoside hydrolases (green), esterase (yellow), hybrid two-component system (HTCS, purple), SusC/SusD-like genes (brown), hypothetical genes (red) and other (gray). (B) SDS-PAGE demonstrating HTCS and scouting enzyme purified close to homogeneity. (C) End-products released from arabinoxylan by the arabinoxylan scouting enzyme after 16 hours of incubation at 37°C. (D) Isothermal titration calorimetry assay determining the binding affinity by the HTCS sensor domain towards the end-products released by the arabinoxylan scouting enzyme.

constituent monosaccharides, i.e., arabinose:xylose and arabinose. During growth on arabinoxylan compared to its monosaccharides, 67 genes were up-regulated at least five-fold in *B. intestinalis.* Out of the 67 up-regulated genes, 26 were predicted to encode carbohydrate related enzymes, including the genes associated with the predicted arabinoxylan-targeting PUL. In fact, the highest up-regulated gene, BACINT_RS19655, is the predicted arabinoxylan-scouting enzyme-encoding gene (Figure 3.3). Previous studies have demonstrated the capacity of the gene products associated with this PUL to fully depolymerize arabinoxylan to its fermentable monomers, i.e., arabinose and xylose, and therefore provide strong support for using the sensor domain of its associated HTCS and scouting enzyme to predict the function of the PUL.

In the experiment with arabinan/sugar beet pulp, *B. intestinalis* up-regulated >57 genes more than 10-fold compared to growth on its constituent monosaccharide arabinose. Our transcriptomic analysis also demonstrated that 42 of the up-regulated genes encode gene products predicted to function as glycoside hydrolases, polysaccharide lyases and carbohydrate esterases. The polypeptides of the up-regulated genes present variable protein domain architectures. The most up-regulated genes occurred in a conserved arabinan utilization loci (PUL44) with associated SusC/SusD pairs, multiple GH families, as well as a hybrid two-component system (Fig 3.4). The other highly expressed genes in this PUL, in all the conditions, encoded enzymes related to arabinose metabolism, highlighting the potential efficiency with which the bacterium uses this cluster for arabinan utilization. Thus, this operon seems to be able to fully depolymerize arabinan substrate and shuttle the monosaccharide into the central metabolism for fermentation.

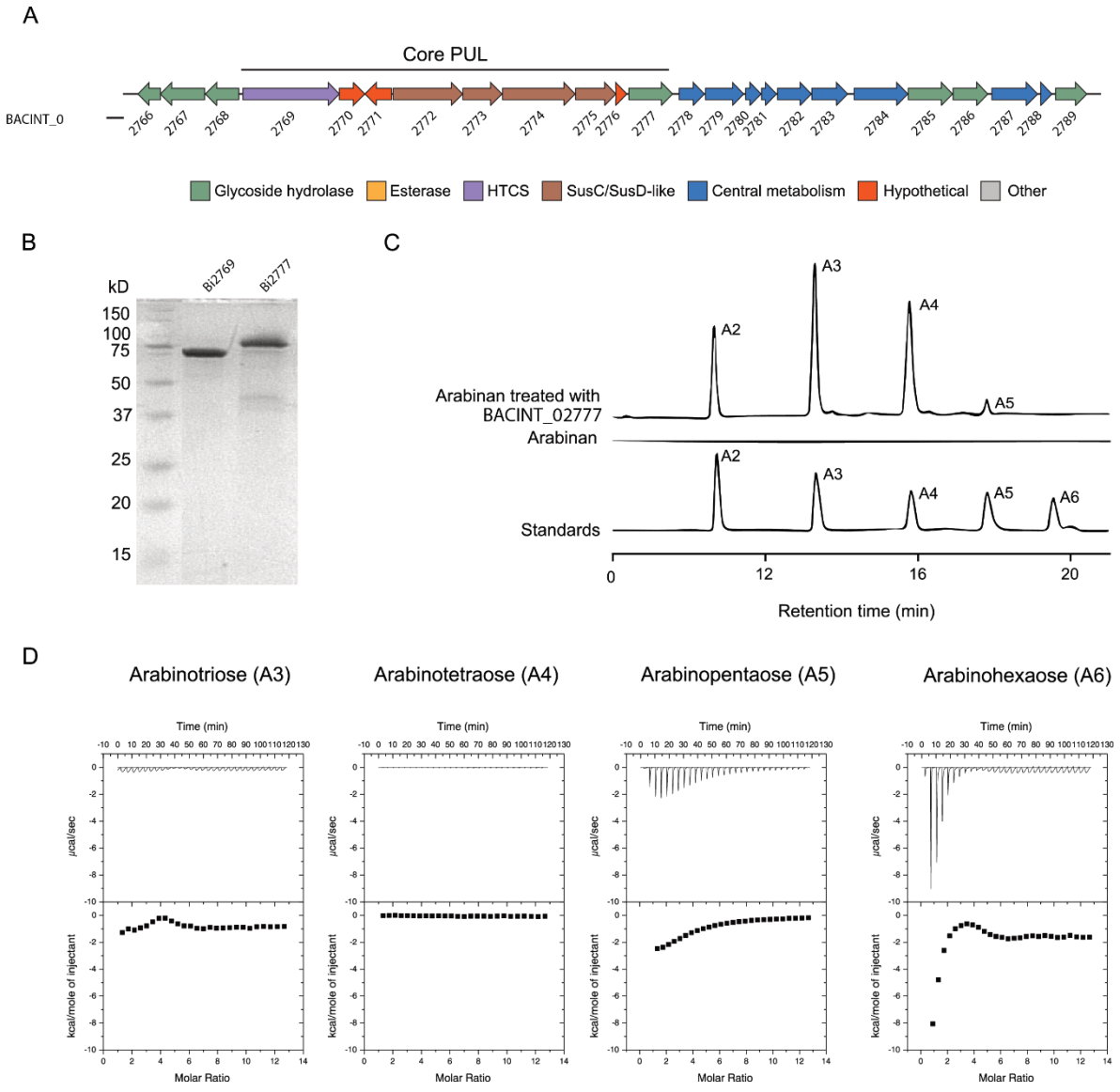


Figure 3.2 - **Identification of an arabinan targeting polysaccharide utilization loci (PUL) in *Bacteroides intestinalis*.** (A) Polysaccharide utilization loci map showing the arabinan-scouting enzyme (BACINT_02777) and Hybrid two-component system (BACINT_02769). PUL is color coded based on predicted functions: Glycoside hydrolases (green), Hybrid two-component system (HTSC, purple), SusC/SusD-like genes (brown), and Hypothetical genes (red). (B) SDS-PAGE analysis of the protein expression of putative endo-α-arabinofuranoside and HTCS sensor domain. (C) End-products released from Arabinan by the arabinan scouting enzyme at 16 hours incubation at 37°C. (D) Isothermal titration calorimetry assay determining the binding affinity by the HTCS sensor domain towards the end-products released by the arabinan scouting enzyme.

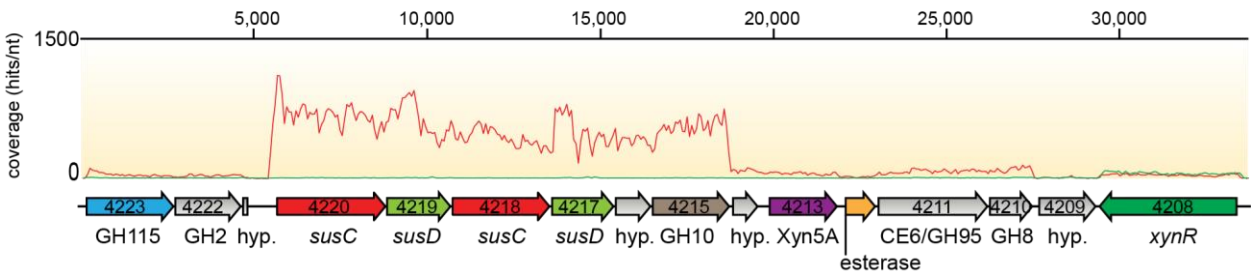


Figure 3.3 - **Trancriptomic analysis of *B. intestinalis* grown in arabinoxylan demonstrates highly expressed arabinoxylan degrading PUL.** Transcriptome map showing RNASeq coverage of the upregulated PUL during growth on arabinoxylan (red) compared to xylose and arabinose mixture (green). PUL is color-coded based on predicted functions: Hybrid two-component system (HTCS, green), SusC/SusD-like genes (red and lime, respectively), and scouting enzyme (brown).

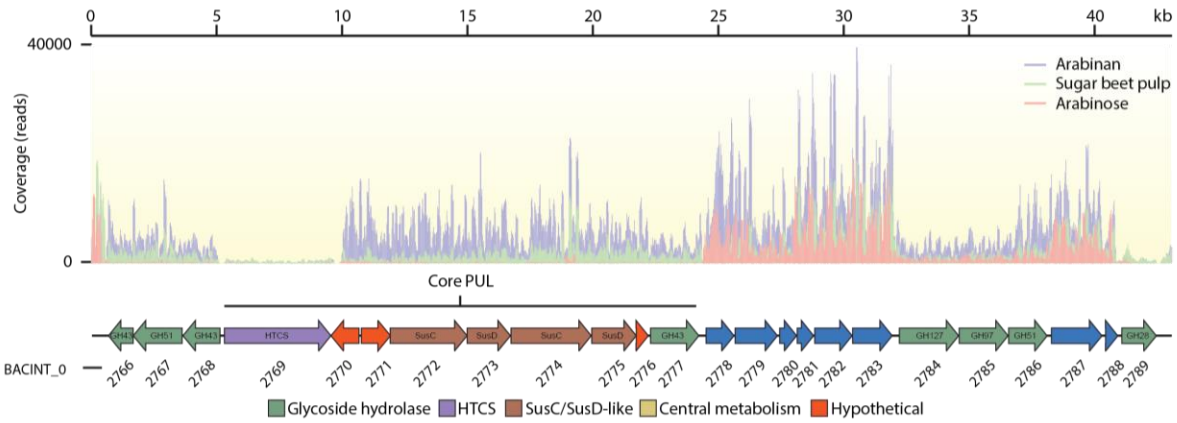


Figure 3.4 - **Trancriptomic analysis of *B. intestinalis* grown in arabinan identifies highly expressed arabinan degrading PUL.** Transcriptome map showing RNASeq coverage of the upregulated PUL during growth on arabinan (purple) and sugar beet pulp (green) compared to arabinose (red). PUL is colored coded based on predicted functions: Hybrid two-component system (HTCS, purple), SusC/SusD-like genes (brown), Central metabolism genes related to the PPP (blue), and Hypothetical genes (red).

Structural analysis of the sensor domain allows prediction of amino acid residues that mediate carbohydrate binding. To provide insights into the mechanism that the hybrid two-component system senses oligosaccharides transported into the periplasm of the microbial cell, we determined the crystal structure of the sensor domain in the apo form. The protein structure is composed of two 7-blade-β-propellers (Figure 3.5), consistent with previous reports. By superimposition of the ligand UA-GlcNAc6S, from the liganded crystal structure of 4A2M, we were able to predict the amino acid residues in the sensor domain within 8 Å distance. There were 16 amino acids within this distance of the ligand, as shown Figure 3.6BC.

Mutational studies of Bi4208 sensor domain sheds light on its carbohydrate sensing mechanism. To shed light on how the sensing mechanism works towards binding xylooligosaccharides, we took advantage of the apo crystal structure of the sensor module of Bi4208 combined with the superimposition of a ligand to predict the amino acids responsible for binding. Based on our analysis, several mutants were created to assess the binding affinity towards xylohexaose. These mutants were expressed and purified close to homogeneity and used in isothermal titration calorimetric analysis to compare their binding to substrate with that of the wild-type polypeptide. Several amino acid residues were determined to play key roles in stabilizing and binding the substrate, and these included a glutamate (E18A), a tryptophan (W462A), and a tyrosine (Y380A) that were each replaced with alanine. Each of the three mutations was able to completely ablate the binding affinity of the sensor domain to xylohexaose (Figure. 3.6A) The results demonstrated these residues as pivotal in coordinating recognition of the substrate by the sensor module of Bi04208.

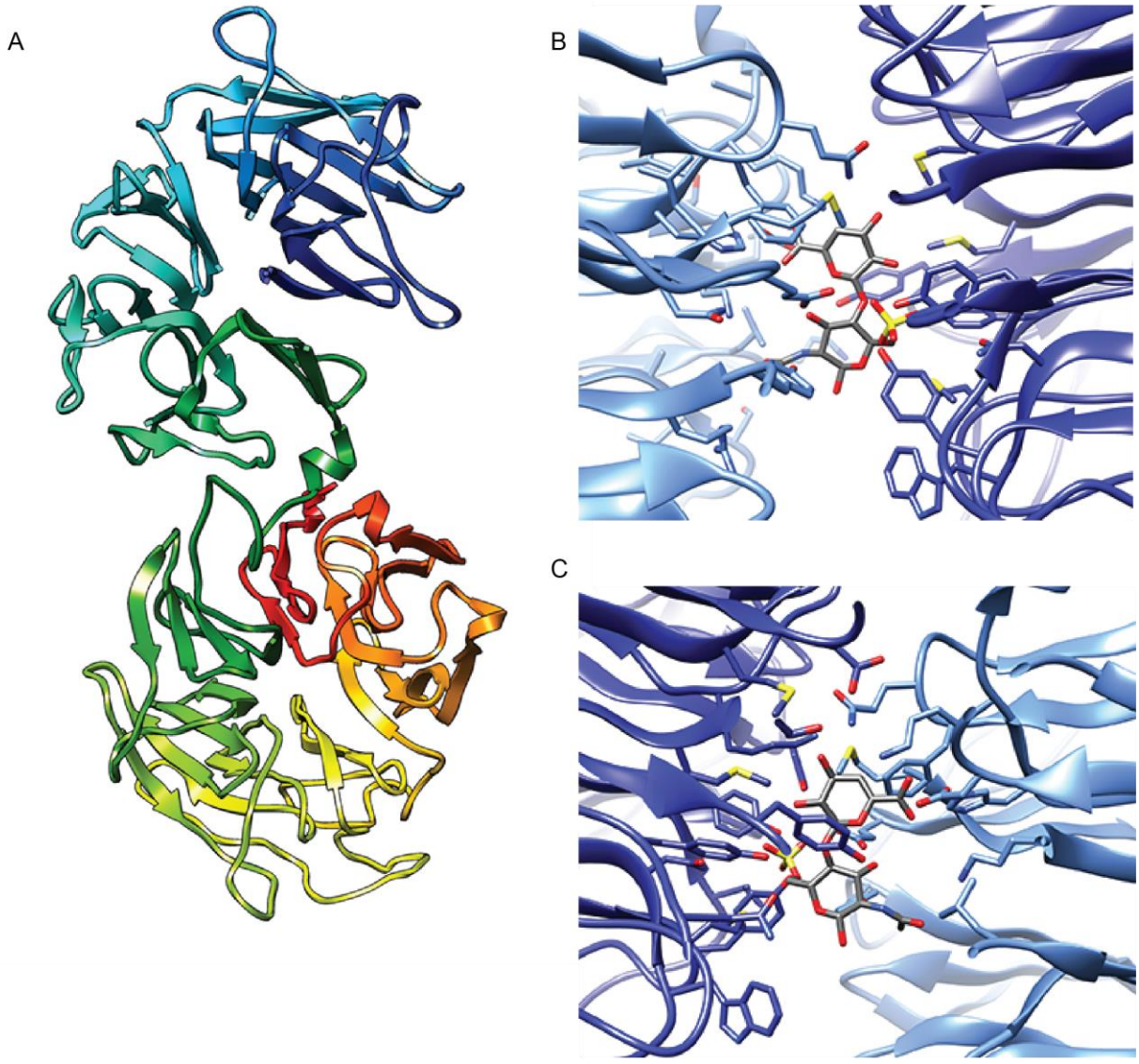


Figure 3.5 - **Overall structure of the arabinoxylan (Bi4208) HTCS sensor domain.** (A) Overall structure of the arabinoxylan HTCS sensor domain demonstrating two 7-β-sheet-propeller. Superimposion of UA-GlcNAc6S demonstrating near amino acids (B and C)

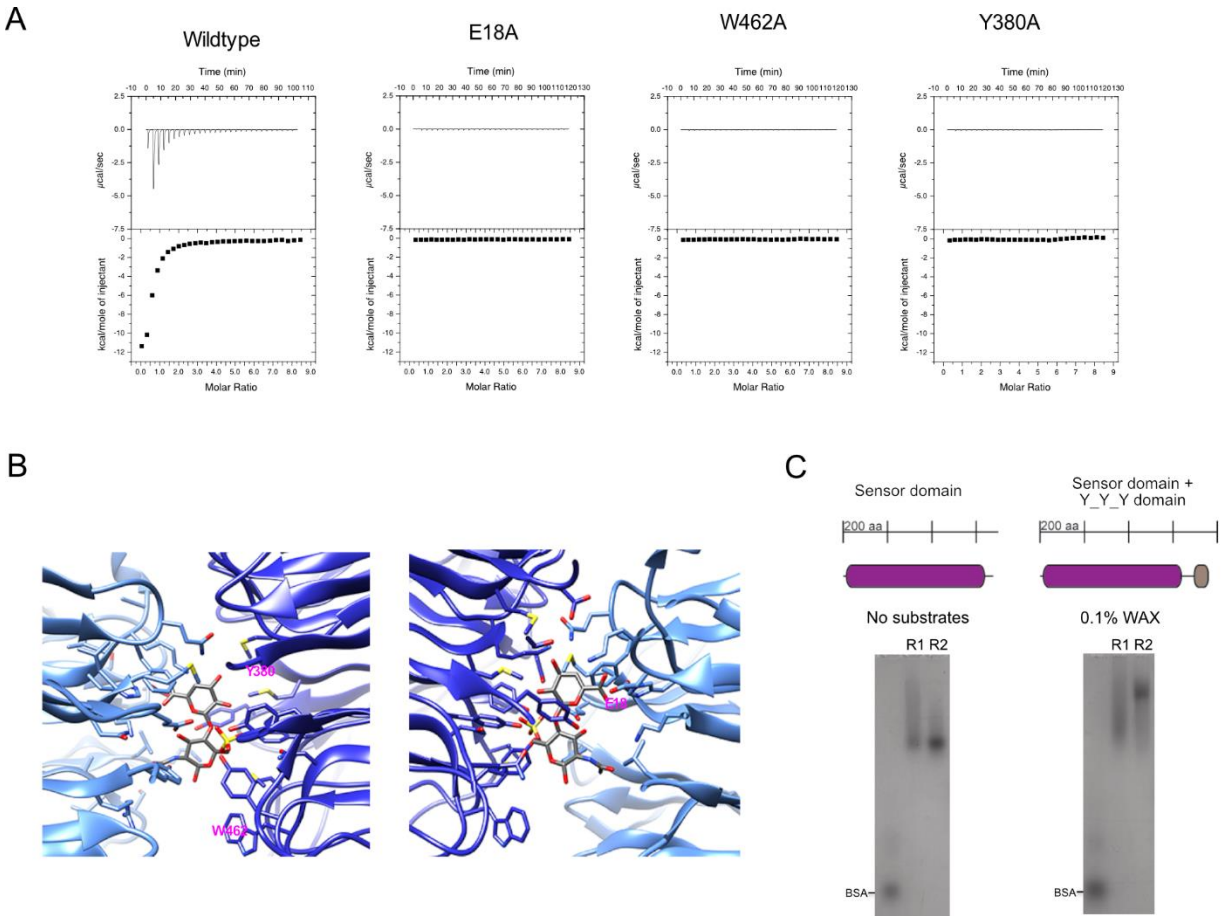


Figure 3.6 - **Mutational studies of the sensor domain of the arabinoxylan HTCS (BACINT_04208).** (A) Isothermal titration calorimetry assay of HTCS wildtype and mutants determining binding affinity by the HTCS. (B) The structure and stereoview of UA-GlcNAc6S superimposion to sensor domain of BACINT_RS19620 demonstrating near amino acids. (C) Gel shift assay in the presence or absence of wheat arabinoxylan of the sensor domain with or without Y_Y_Y domain.

It was also the Y_Y_Y domain seems to play an important role in stabilizing the protein to allow binding. This domain contains three conserved tyrosine residues near the sensor domain that we predict help to engage the substrate. Currently there is no known function described for the Y_Y_Y domain, although this module is highly conserved in the HTCS family of proteins. As preliminary test for its importance for binding, we assess its function by affinity native gel electrophoresis, and based on the results, it appears that the Y_Y_Y domain plays a key role in binding of the substrate (Figure 3.6C).

Arabinoxylan hybrid two-component system is conserved in different *Bacteroides* species. Arabinoxylan degradation is a key process for several *Bacteroides* species. Previous reports have demonstrated that the arabinoxylan PUL from *B. intestinalis* is conserved in several members of the Bacteroidetes phylum, including organisms from soil, rumen, and human gut environments. To understand the conservation of this sensor domain in other species, we perform a multiple sequence alignment of HTCS sensor domain of key human gut species. The results demonstrated the conservation of this sensor domain in other *Bacteroides* as depicted by Figure 3.7. These proteins exhibited very high conservation and portrayed identities from 59.08% to 82.70% (Table 3.1). Interestingly, the HTCS in three species, *B. oleiciplenus*, *B. intestinalis*, and *B. cellulosilyticus*, appeared to be part of a similar cluster, while in *B. eggerthii* the cluster seems to have undergone some modification (Figure 3.8).

HTCS sensor domain is membrane bound in *Bacteroides* species. The canonical two component system suggests that the sensor domain is associated with the periplasmic region, while the response regulator is in the cytoplasm. Due to the fusion of

Figure 3.7 - **Alignment of arabinoxylan related HTCS sensor domain among *Bacteroides* species demonstrate high conservation of the polypeptide sequences.** Multiple sequence alignment of proteins encoded by BACINT_04208 (*B. intestinalis*), BACEGG_01311 (B. eggerthii), HMPREF9447_0146 (*B. oleiciplenus*), and BACCELL_03420 (*B. cellulosilyticus*) demonstrates high conservation.

Table 3.1 – Identity matrix (%) between arabinoxylan HTCS sensor domain of different *Bacteroides* species.

	B. eggerthii	*B. oleiciplenus*	*B. intestinalis*	*B. cellulosilyticus*
B. eggerthii		57.6	59.0	58.7
B. oleiciplenus	57.6		69.4	67.2
B. intestinalis	59.0	69.4		81.4
B. cellulosilyticus	58.7	67.2	81.4	

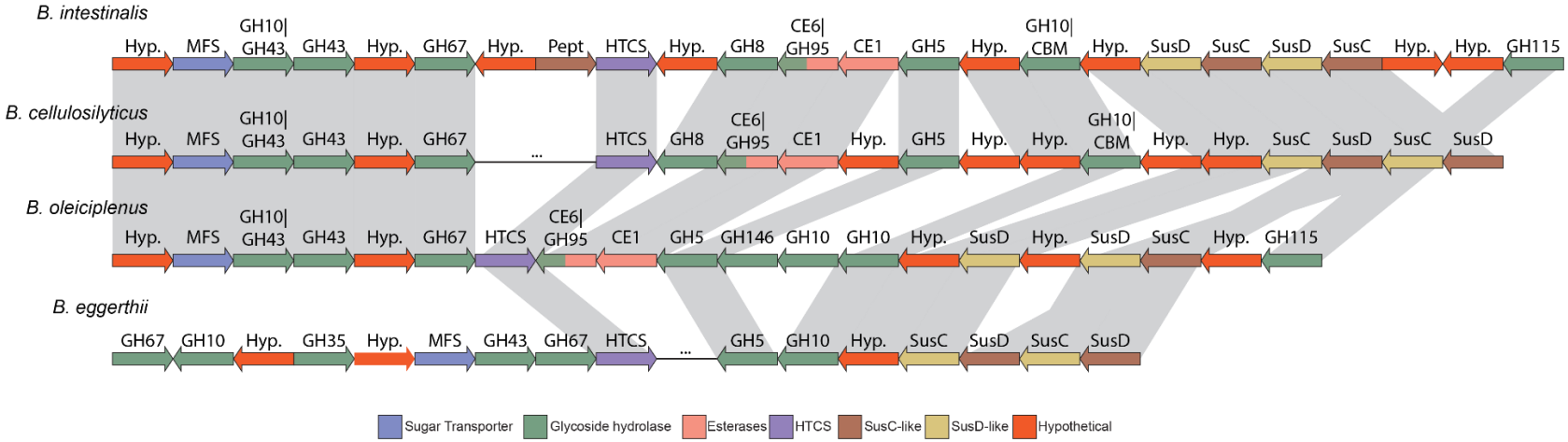


Figure 3.8 - **Genomic context and multiple genome alignment demonstrate relationship of arabinoxylan PUL in closely related species.** Representation of arabinoxylan degrading PUL in primary degrader *Bacteroides* species. PUL core genes are present in all species (SusC/D, HTCS, and scouting enzyme (GH10)

the two genes into one coding for a single polypeptide, we sought to understand the localization of the sensor domain of the HTCS in different *Bacteroides* species. For this, purpose, we developed polyclonal antibodies against the sensor module of the arabinoxylan HTCS of *Bacteroides intestinalis*. Furthermore, due to its conservation with other members of the colonic environment, we hypothesized that the antibody will cross-react with its homologs in different species. We cultured the bacteria in minimal medium containing arabinoxylan as a sole carbon source and fixed them for immunostaining. As demonstrated in the images, the HTCS sensor domain is indeed associated with the membrane in all four Bacteroides species tested in this study, and hence demonstrates conservation in the localization of this polypeptide in these bacteria (Figure 3.9).

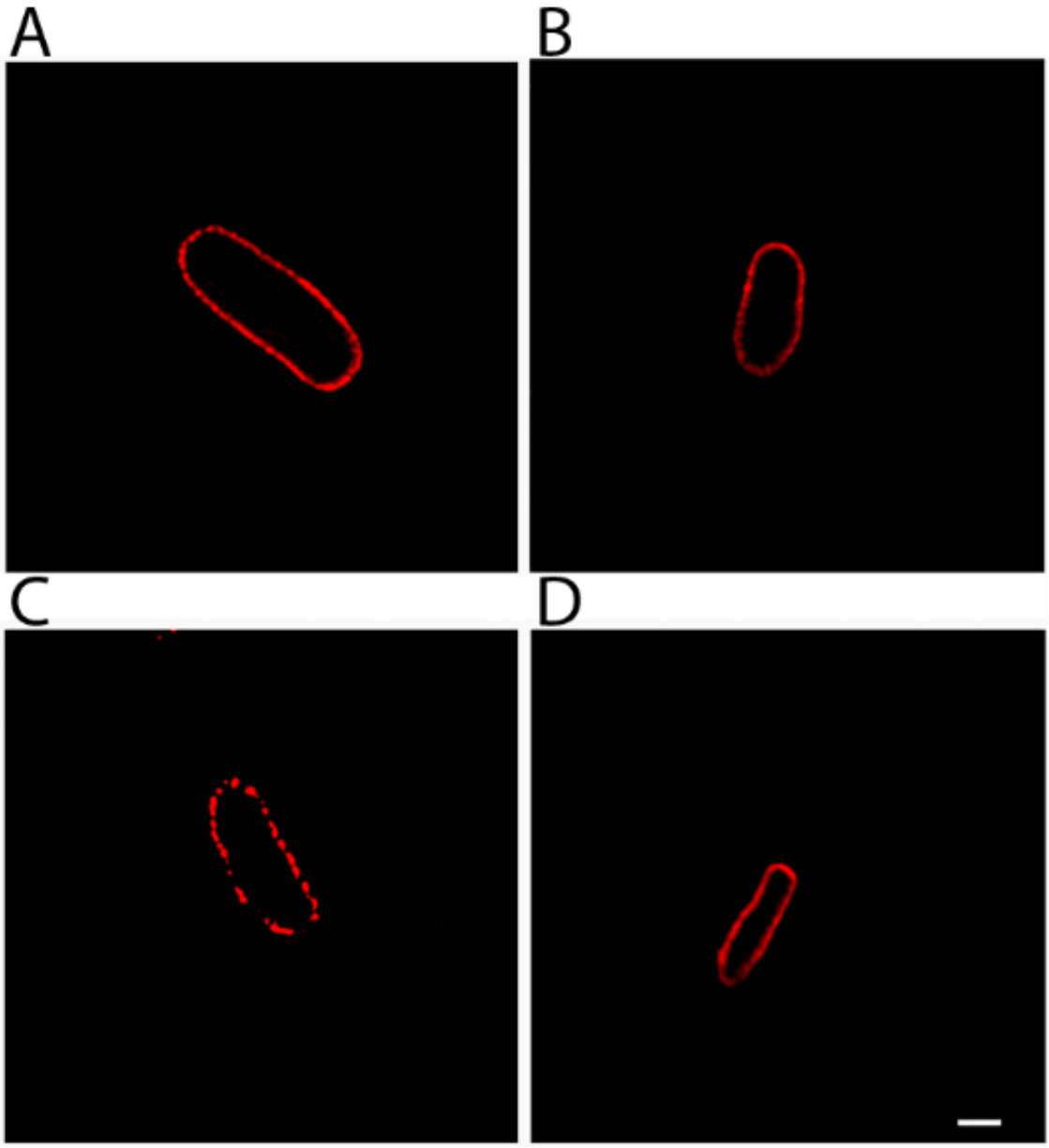


Figure 3.9 - **Immunostaining of arabinoxylan HTCS sensor domain demonstrates membrane localization in different *Bacteroides* species.** Superresolution structured illuminarion microscopy (SR-SIM) of HTCS sensor domain in *Bacteroides oleiciplenus* (A), *Bacteroides* eggerthii (B), *Bacteroides cellulosilyticus* (C), and *Bacteroides intestinalis* (D), using α-Bi4208 sensor domain. These results were obtained in collaboration with Dr. Kingsley Boateng from UIUC Carl R. Woese Institute for Genomic Biology Core Facilities.

3.3 Discussion

Bacteroidetes possess a remarkable array of polysaccharide degrading enzymes organized in operons termed polysaccharide utilization loci (Martens, Chiang et al. 2008, Zhang, Chekan et al. 2014). Currently, there is about 5,000 different PUL deposited at the PUL database (http://www.cazy.org/PULDB/), with the vast majority without associated function. In general, to assign polysaccharide targets for a PUL, a biochemical approach or transcriptomic analysis that identifies the function of enzymes and how this PUL is regulated *in vivo* is applied. In this study, we aim to develop a more rapid approach to identify the potential targets of the large number of uncharacterized PULs. The canonical two-component system are the dominant transmembrane mechanism to signal environment changes in bacteria and regulate key biological changes in response (Hoch and Silhavy 1995). The hybrid two-component system in Bacteroidetes seem to act in a similar way, since they appear to sense oligosaccharides and activate gene transcription of their cognate PUL (Lowe, Baslé et al. 2012). However, the mechanism of signal transfer between the membrane bound module and the cytoplasmic domain remain elusive. In the present studies, we provide evidence that a key commonality of the hybrid two-component system with the classic two component system is the localization of the sensor module in the periplasmic region. By inference and preliminary studies in our lab, we predict that the response regulators of these polypeptides are located in the cytoplasmic region.

Recent studies have demonstrated that *susC/susD*-like genes are the key genes expressed in different PULs (Martens, Lowe et al. 2011, McNulty, Wu et al. 2013, Zhang, Chekan et al. 2014, Despres, Forano et al. 2016). Thus, comparison of different PULs,

demonstrated that several PULs possess a potentially endo-acting GH following the SusC/D pair. Therefore, we hypothesized that this enzyme may act as a scouting enzyme during low expression of the PUL. Low-level expression of the endo-acting enzyme leads to cleavage of the cognate polysaccharide, upon encounter in the organisms environment, into oligosaccharides that are then sensed by the HTCS sensor domain present in the periplasmic region. In other words, this enzyme “scouts” the environment for its cognate polysaccharide, allowing the HTCS sensor domain to sense and activate the expression of the degradative enzymes encoded in the gene cluster or PUL. Taking advantage of this substrate specificity of the sensor domain region combined with the activities of the scouting enzymes, we aim to develop a more rapid approach to assign function to uncharacterized PULs. Therefore, to test this approach, we expressed and purified two different scouting enzymes from a predicted arabinoxylan and arabinan PULs. In fact, both enzymes were able to recognize their predicted polysaccharide substrates and degraded them into oligosaccharides. The arabinoxylan-scouting enzyme released products ranging from xylose to xylopentaose (Figure 3.1C), while the arabinan-scouting enzyme cleaved the polysaccharide into oligosaccharides ranging from arabinobiose to arabinohexaose (Figure 3.2C). To further probe this hypothesis, we performed isothermal titration calorimetric binding assays with the purified HTCS sensor domain of each PUL to determine whether they bind to the end products of the cognate scouting enzyme. The arabinoxylan sensor domains bind to the larger oligosaccharides, xylopentaose and xylohexaose (Figure 3.1D), while the arabinan sensor domain demonstrate higher affinity towards arabinohexaose (Figure 3.2D). We hypothesized that, although the scouting enzymes release smaller oligos, during our *in vitro* experiments,

larger oligosaccharides are likely initially released *in vivo* and sensed by the HTCS. Thus, the activity of the scouting enzymes is able to release the oligosaccharides sensed by the HTCS sensor domain. These results provide a new approach to address or annotate the large diversity of PULs reported for the Bacteroidetes. In other words, by identifying and characterizing the scouting enzyme and the sensor module of the associated HTCS, the function of the gene cluster can be assigned. Furthermore, transcriptomic analysis of *B. intestinalis* grown in the substrates tested, i.e., arabinoxylan and arabinan, compared to their monomeric units confirmed the utility of our proposed approach to assigning function to a PUL (Figure 3.3 and 3.4, respectively).

The arabinoxylan HTCS encoded by BACINT_04208 is composed of two 7-blade-β-propellers and superimposition of the ligand (UA-GlcNAc6S) from BT4663 sensor domain allow us to find candidates for the amino acid residues responsible for binding to substrate. Interestingly, mutagenesis of amino acids from both propellers are able to ablate the binding affinity of this protein for xylohexaose (Figure 3.6A), demonstrating that the protein requires both propellers to bind the substrate. Recent structural studies have demonstrated that HTCS for a homo-dimer. Thus, the sensor domain may form an asymmetric unit, utilizing the N-terminus and C-terminus protomer from different chains to stabilize the substrate binding. Furthermore, the Y_Y_Y domain seems to be a key domain for the binding of the sensor domain. We demonstrated that a truncated mutant without the Y_Y_Y domain is not able to bind the soluble wheat arabinoxylan, while the sensor domain together with the Y_Y_Y domain binds to the substrate. Although the function of Y_Y_Y domain is not completely understood, this conserved region seems to be involved in the dimerization of the sensor domain (Lowe, Baslé et al. 2012).

Furthermore, this HTCS is conserved in several known arabinoxylan degraders, such as *B. cellulosilyticus, B. oleiciplenus,* and *B. eggerthii*. Thus, these bacteria seem to employ a similar sensing mechanism as *B. intestinalis*. Despite small differences in the clusters in these bacteria, the core arabinoxylan degrading enzymes composed of the SusC/SusD, HTCS, and scouting enzyme are conserved. Interestingly, *B. cellulosilyticus* and *B. eggerthii* have an insertion inside the cluster as depicted by Figure 3.8.

The current model for HTCS is that the sensor domain is associated with the membrane, spanning the response regulator domains to the cytoplasmic region (Lowe, Baslé et al. 2012, Lynch and Sonnenburg 2012). To address this hypothesis, we developed antibodies against the arabinoxylan HTCS sensor domain encoded by BACINT_04208 and assessed its localization. The results suggested that the sensor domains of the arabinoxylan HTCS of the four arabinoxylan degrading bacteria are associated with the membrane, and therefore corroborating the hypothesis.

The results demonstrated in this study may advance the understanding of polysaccharide utilization by Bacteroidetes species by rapidly assigning function to previously uncharacterized PULs. Our findings suggest specificity of the sensor modules for the end products of hydrolysis of the scouting enzyme, and thus allow determination of the potential PUL target. This process leads to a faster approach to identify diverse PULs functions. Further studies may allow us to increase the understanding of each particular polysaccharide utilization system, generating a signature for polysaccharide sensing. The procedures outlined in this work and the results can be used to analyze metagenomics samples of individuals to identify their polysaccharide degradation

signature, which can be used for targeted dietary interventions in order to give a competitive advantage to commensal/beneficial Bacteroidetes.

3.4 Materials and methods

Materials. The human colonic bacterial isolates *Bacteroides intestinalis* DSM 17393 was used in this study. *E. coli* DHα and BL21-CodonPlus was used as competent cells. Phusion High-Fideltiy DNA Polymerase and all required restriction enzymes were obtained from New England BioLabs. The pET-28 and pET-46 was obtained from Novagen. QIAprep Spin Miniprep, PCR purification, RNAprotect, RNase-free DNAase and RNeasy kits were obtained from Qiagen. The Talen Metal Affinity resin was obtained from Clontech Laboratories, Inc. The xylooligosaccharides, arabinan, etc. was obtained from Megazyme. Pectin, arabinoxylan (beechwood) was obtained for Sigma-Aldrich.

Growth of *Bacteroides* spp.. *Bacteroides* strains were routinely cultured anaerobically at 37ºC in Brain heart infusion (BHI) with Hemin and cysteine. When necessary, cells were inoculated into defined media with appropriate carbon sources. All media compositions are listed in Table 1 and Table 2. The optical density at 600 nm wavelength was monitored using a Spectronic 20D (ThermoFisher Scientific) over a period of 27 h. At mid-late exponential phase of growth (OD_{600} = 0.4), 50 mL of each culture was removed, combined with two volumes of RNAprotect, and RNA was extracted using the RNeasy mini kit based on the manufacturer's protocol. The optional on-column DNase treatment was performed to ensure total removal of DNA. The RNA was then quantitated using a Qubit RNA assay kit (Invitrogen) and stored at −80 °C until RNA sequencing as described in our earlier report (22). RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara CA) with all samples achieving a RNA

Integrity Number (RIN) of 9 or higher. The High-Throughput Sequencing and Genotyping Unit of the Roy J. Carver Biotechnology Center at the University of Illinois Urbana Champaign performed high throughput sequencing using an Illumina 4000 sequencing platform.

Resulting sequence data was analyzed for differentially expressed genes following a previously published protocol (Anders, McCarthy et al. 2013). Briefly, reads were filtered for quality using Trimmomatic (Bolger, Lohse et al. 2014). Reads were aligned to the genome using BowTie2 (Langmead and Salzberg 2012). Reads mapping to gene features were counted using htseq-count (Anders, Pyl et al. 2015). Differential expression analysis was performed using the edgeR package in R and the TMM method was used for library normalization (Robinson, McCarthy et al. 2010). Coverage data was visualized using Integrated Genome Viewer (IGV) (Thorvaldsdottir, Robinson et al. 2013).

Gene cloning, site-directed mutagenesis and expression. All genes cloned for expression were amplified from *B. intestinalis* DSM 17393 genomic DNA by PCR using the primers listed in Table 3.1. The cloning of the correct PCR products into pET- 46b vector, overexpression of the genes, and site-directed mutagenesis were as described in our earlier report (Yoshida, Mackie et al. 2010). The recombinant proteins in the cell lysates were purified with Talon metal affinity resin (red), and the purity was assessed by 12% SDS-PAGE with Laemmli's method (Laemmli 1970). The protein concentrations were determined by the method of Gill and von Hippel based on the molecular mass and computed extinction coefficients (Gill and Von Hippel 1989).

Purification of the recombinant proteins The expressed cells were harvested

by centrifugation (4,000 × *g*, 4°C) for 30 min and resuspended in lysis buffer (30 ml; 50 mM Tris-HCl, 300 mM NaCl [pH 7.0]). The cells were then lysed by three passages through an EmulsiFlex C-3 cell homogenizer from Avestin (Ottawa, Canada). The cell debris/insoluble fractions was removed by centrifugation (20,000 × *g*, 4°C) for 20 min. The soluble fraction that contains hexahistine tag was further purified via immobilized metal affinity chromatography using Talon metal affinity resin (Clontech, Mountain View, CA). After resin was equilibrated with binding buffer (50 mM Tris-HCl, 300 mM NaCl [pH 7.5]) solution fraction were added and incubated for 1 h at 4°C with gentle rocking. The resin slurry was then transferred to a gravity column, and after washing with 50 column volumes of binding buffer and washing buffer (binding buffer with 10 mM imidazole [pH 7.5])., the bound protein was eluted with 10 column volumes of elution buffer (binding buffer with 250 mM imidazole [pH 7.5]).

If necessary for further purification, the elution fractions of each of the proteins were pooled dialyzed to exchange into an anion-exchange column binding buffer (20 mM Tris-HCl, pH 7.6). The proteins were concentrated with Amicon Ultra-15 centrifugal filter units (50,000-Da molecular separately mass cutoff) and loaded onto a 5-ml HiTrap Q anion-exchange column (GE Healthcare, Piscataway, NJ), and eluted with a linear gradient of elution buffer (20 mM; 1 M NaCl, pH 7.6) from 0 to 50% over 20 column volumes and then from 50 to 100% over 5 column volumes. The fractions containing the purified protein were pooled and further purified with a Superdex 200 Hiload 16/60 size exclusion column with protein storage buffer (50 mM Tris-HCl, 150 mM NaCl [pH 7.5]) as the mobile phase. The anion-exchange and gel filtration chromatography assays were performed using an AKTAxpress fast protein liquid chromatograph equipment (GE

Healthcare, Piscataway, NJ). The collected fractions were pooled, concentrated, and analyzed with 12% SDS-PAGE.

Hydrolysis and detection of oligosaccharides. Hydrolysis of oligosaccharides and the natural polysaccharide substrates by various scouting enzymes were performed at 37 °C for 16 h in citrate buffer (50 mM citrate, 150 mM NaCl, pH 5.5). The end products of hydrolysis were analyzed by reducing sugar assays, and high-performance anion exchange chromatography (HPAEC), all as described previously (Wefers, Cavalcante et al. 2017). Briefly, The end products were analyzed by HPAEC-PAD on an ICS-5000 system (Thermo Scientific Dionex, Sunnyvale, CA) equipped with a CarboPac PA-100 column (250 mm × 2 mm; Thermo Scientific Dionex). The flow rate was 0.25 mL/min, and a gradient composed of the following eluents were used at 25 °C: (A) 0.1 M sodium hydroxide and (B) 0.1 M sodium hydroxide + 0.1 M sodium acetate.

Affinity gel electrophoretic mobility assays. Assay to monitor the binding affinity of Bi4208 sensor domain with and without Y_Y_Y domain were performed according to previous reports. In short, the stacking gel contained 3% wt/vol polyacrylamide in 1.5M Tris-HCl buffer (pH 8.3). Each protein was loaded in a native gel (12% wt/vol polyacrylamide) with or without soluble wheat arabinoxylan (0.1% wt/vol), and the gels were electropheresed simultaneously at 4°C at 100V for 6h. The gels were then stained with Coomassie Brilliant Blue G-250 for protein visualization.

HTCS sensor domain binding assays. The Isothermal titration calorimetric (ITC) measurements were performed using a VP-ITC microcalorimeter with a 1.4 mL cell volume from MicroCal, Inc. The proteins were exchanged into phosphate buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.0) by dialysis, and the oligosaccharide ligands

Table 3.2 – Primers used in this study.

Primer (5'-3')	Gene	Sequence[a]
Bi4208For	BACINT_04208	GACGACGACAAGATGCAAGAAGGAAAAATGTTCACAGTGGATAGAGAATTATCAAGC
Bi4208Rev		GAGGAGAAGCCCGGTAACCATGCAGGATGGATTGTGATACTAAATTCTTTGATGTCTG
Bi4215For	BACINT_04215	GACGACGACAAGATGAAATATACGAAAAATATCTTAGGAACAATGCTTTTAATGG
Bi4215Rev		GAGGAGAAGCCCGGTTATTTTCCAGCCAATCCATCAGCAAAAC
Bi2769For	BACINT_02769	GACGACGACAAGATGCAAGAACATTTTGCAGATCGATACAACGTTTC
Bi2769Rev		GAGGAGAAGCCCGGTTAGGGTAGGATGGTGATTTTTAGCTCTGC
Bi2777For	BACINT_02777	GACGACGACAAGATGAGCGAAAGCAATGAGCCTGACATAG
Bi2777Rev		GAGGAGAAGCCCGGTTATTTCTTCTTTTTACCCCAATACGTCTTGCTGTTG
Bi4208D239AFor	BACINT_04208	GAATCAAATTCTTATTGGTACAGCGGGTTCGGGAATGAAAGTTTATG
Bi4208E18AFor	BACINT_04208	AATCATCTGGATAGCAACAGCGGATGGTCTGAATCGATACGA
Bi4208F286AFor	BACINT_04208	GAAATATATGGTTAGGGTTCGCGCAGAAAGGAGTTATGATTAT
Bi4208K224AFor	BACINT_04208	CTAATCCCAACTTACCTATCGCGATCCTATACCCTGAGGAGGA
Bi4208K269AFor	BACINT_04208	CATTTGACCTTACAAAAGCAGCGATACATTCTATTCTGAAAG
Bi4208Q287AFor	BACINT_04208	GAAATATATGGTTAGGGTTCTTTGCGAAAGGAGTTATGATTATACCTG
Bi4208W462AFor	BACINT_04208	GTAAATATGCTACATAACTCAGCGATAAATTGCCTGTTACATAC
Bi4208Y169A For	BACINT_04208	ACAACAATCAATTCAAACACGCGTCCGGCGAAAAAAGCATAGC
Bi4208Y305AFor	BACINT_04208	ACGAGTTCAAGTATCTAGGTGCCAAATCTTCCAGACACAATG
Bi4208Y380AFor	BACINT_04208	ACAATATGTGGATAGGTTCTGCGTTGAATGGCATGGCCCGATT
Bi4208Y411AFor	BACINT_04208	GAAATATATGGTTAGGGTTCGCGCAGAAAGGAGTTATGATTAT
Bi4208Y46AFor	BACINT_04208	CGAATTCCTTATTAAGTGACGCGGTACGTACACTGTTCGAAGA
Bi4208Y479AFor	BACINT_04208	CGCTGAAGACAAACTCTACATAGGTACTGCGGATGGTTTGGGCTGTCTGGATATAAAAAC
Bi4208Y508AFor	BACINT_04208	GAATATTATATGGCGATGTAATCGCGGCTTTGCATGAAGGTAAGGATG
Bi4208Y567AFor	BACINT_04208	ATTTTTGGATCAGTACGAATGCGGGAATCAGCCGATTCAATCC
Bi4208Y583AFor	BACINT_04208	ACCAAAGTTTCATTAATTTTGCGGCAAGTGACGGATTACAAAG

were dissolved in the same buffer. The proteins (50 μM) were then injected with 28 successive 10 μL aliquots of ligand (2 mM) at 300-s intervals. The data were fitted to a nonlinear regression model using a single binding site (MicroCal Origin software). The thermodynamic parameters were calculated using the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$), and the relationship $\Delta G = -RT\ln Ka$.

Immunostaining and microscope pictures. *Bacteroides* cells cultured under arabinose/xylose and wheat arabinoxylan were harvested during mid-log phase for immunostaining. In short, one mL of cells were centrifuged (4.000g for 5 minutes) and washed twice in PBS before being re-suspended in one ml of PBS. The cell suspension was fixed by mixing 1:3 v/v in 4% paraformaldehyde in PBS for 4 hours. After fixing, 20 μL of cells were spread onto 20x30mm cover slip and incubated in a moist chamber for an hour. The cells in the coverslip were washed in PBS for 5 minutes. Coverslips were incubated with blocking buffer (2% BSA, 1% goat serum, 0.05% Tween-20, in PBS) for 30 minutes at room temperature. The coverslips were briefly washed with PBS and incubated with 1:500 BACINT_04208 polyclonal antibody in blocking buffer for 2 hours. After primary antibody incubation, coverslips were washed three times in PBS for 5 minutes each and incubated with 1:1000 diluted secondary antibody (Cy3-conjugated anti-rabbit) in blocking buffer. The coverslips were washed in PBS three times and rinsed one minute in 0.04% photo-fluo. The coverslips were inverted onto microscope slides spoted with 25 μL Vectra Shield (Vector Co.) containing 1.5 μL of DAPI. The slides were observed with confocal microscope fitted with 100X oil immersion lens. The images were captured, analyzed with ZEN software, and three-dimensional representations were

constructed using the blind 3D deconvolution function of the software from 10 Z-stacked images taken at 0.1 µm increments.

Crystallography Studies. BACINT_04208 was crystallized by screening with sparse matrix conditions followed by optimization. Specifically, BACINT_04208 crystals were optimized to 20 mg/mL in 50mM Tris-HCl, pH 7.5, and 150mM NaCl. The hanging drops were set at 4°C using 2 µL of protein to 2 µL of a stabilizing solution of 15% polyethylene glycol, 0.2M lithium sulfate monohydrate, and 0.1 M Tris-HCl pH 8.5. Crystals were soaked in mother liquor containing 30% (wt/vol) threitol before vitrification by direct immersion into liquid nitrogen. X-ray diffraction data of BACINT_04208 crystals were collected at the Life Science Collaborative Access Team, Sector 21, Argonne National Laboratory.

Crystallographic phases for BACINT_04208 was determined by anomalous diffraction data collected from crystals soaked with 1 mM 4-chloromercuribenzoic acid for 4 h at 23 °C. After soaking, 4-chloromercuribenzoic acid-derived crystals were treated identically as native crystals. Data were indexed and scaled using HKL2000 or autoPROC. Phasing was carried out using the AutoSol package from the software PHENIX suite. Solutions from AutoSol were automatically built using ARP/wARP. The structures were further refined using a combination of COOT, REFMAC5, and PHENIX refine.

CHAPTER 4: CONCLUSION AND FUTURE PROSPECTIVE

The Bacteroidetes possess one of the most extensive polysaccharide utilization systems known in bacteria and despite intensive research over the last 30 years, most of these loci remain largely uncharacterized. As the Bacteroidetes are one of the dominant phyla in the human gut, understanding their impact on the nutritional and metabolic potential within this environment may increase the capacity to modulate host physiology. Furthermore, the utilization of complex esterified substrates by this phylum has not been extensively studied. Due to recent interest in the metabolic potential of the microbiome and the benefits of phenolic compounds released in the gastrointestinal tract, understanding the utilization of esterified substrates has become a key area of research. Esterified polysaccharides are largely present in dietary cereals such as wheat, rice, and oats. In this study, we sought to further expand the understanding of polysaccharide utilization by the Bacteroidetes through comparative genomics, whole genome transcriptomics, biochemical and structural analysis. Through our work, we describe the regulatory and biochemical mechanisms that various *Bacteroides* spp. employ to fully degrade highly esterified arabinoxylan under different conditions.

Our findings demonstrate that, despite encoding a large PUL that targets arabinoxylan, these bacteria employ a different strategy in degrading highly esterified substrates of similar structural backbone. These results underscore the importance of arabinoxylan for the physiology of the Bacteroidetes primary degraders and consequently for the host. Enzymes encoded by the esterase-enriched cluster are able to work synergistically to degrade highly esterified substrates such as wheat bran. Furthermore, biochemical analyses of an uncharacterized hypothetical protein from *B. intestinalis*

identified a novel, versatile, feruloyl esterase capable of cleaving the ferulic acid side chain from all the tested substrates. In fact, through sequence analysis of all known carbohydrate esterases encoded by Bacteroidetes species we established that this hypothetical protein is a member of a novel class of carbohydrate esterases distinct from all previously classified families. Interestingly, phenolic compounds released by the *Bacteroides* through the action of these esterases were not further metabolized by the species encoding the esterases. Thus, the function of this esterase-enriched cluster may be to provide ferulic acid to the host or other members of the gut community. Based on our results, ferulic acid released by these esterases can increase the capacity of the host to respond to intracellular pathogens through modulation of the immune response.

Uncovering the mechanism of highly esterified substrates by *Bacteroides* spp. provides a starting point for future research. Throughout this work, many questions have remained elusive, e.g. the polysaccharide preference of these bacteria. Our work and previous research has demonstrated the ability of *B. intestinalis* to degrade both soluble wheat arabinoxylan as well as the highly esterified substrate through two distinct clusters. Thus, a next line of experiments may aim to understand the utilization and preference of these similar, yet distinct, substrates *in vivo*. Illuminating the mechanisms employed by these bacteria within the host setting may provide answers regarding its metabolic impact in the human gut. Furthermore, although we demonstrated the degradation of esterified substrates and ferulic acid accumulation during growth of *Bacteroides* spp. in the complex wheat arabinoxylan substrate, it is important to note that these *in vitro* experiments may not reflect how these substrates are actually metabolized *in vivo*. To further understand its *in vivo* impact, mono-colonization of mice with different arabinoxylan-degrading

species, in an esterified arabinoxylan-rich diet, may allow better understanding of the impact of the bacterial metabolism towards the host. Furthermore, these experiments may shed light on the amount of ferulic acid released in a more complex environment, as well as its absorption and secretion by the host. The ultimate goal, is to understand how the primary arabinoxylan-degraders behave and interact with the host/microbiome community. Defining the metabolism of ferulic acid and its derivatives by the microbial community will improve our ability to modulate this metabolic function to benefit the host.

Due to the vast number of uncharacterized PULs encoded in the microbiome, it is important to develop a rapid means to assign function to these gene clusters. Our current hypothesis relies on the observation of a potential endo-acting enzyme near the *susC*/*susD*-like genes. Previous transcriptomic analysis shows that this potential scouting enzyme is highly expressed during growth in the presence of its cognate polysaccharide (McNulty, Wu et al. 2013, Zhang, Chekan et al. 2014, Despres, Forano et al. 2016). However, these enzymes appear to be expressed at low-levels even in the absence of the target polysaccharide, suggesting that these enzymes “scout” the environment for its associated target polysaccharides. In order to probe this hypothesis, in this work, we biochemically characterized the hydrolytic activities of two scouting enzymes on arabinoxylan and arabinan. Furthermore, we show that a PUL-associated HTCS sensor domain is able to bind and sense the products of its associated scouting enzyme. In doing so, we demonstrate that the scouting enzyme and HTCS may work in combination to regulate the expression of the gene cluster. This hypothesis was further corroborated by transcriptomic analysis of *B. intestinalis* under growth in the polysaccharide compared to its monomeric components. To further shed light on the sensing mechanism of the HTCS

sensor domain, we solve the structure of this domain encoded by BACINT_04208. Interestingly, this domain is composed of two 7-β-sheet-propellers that seem to be critical for substrate recognition. Moreover, this mechanism is conserved in several arabinoxylan-degrading *Bacteroides* spp.

While polysaccharide utilization by the Bacteroidetes has been extensively studied, the vast diversity of PUL activities demonstrates the complexity of these systems. Although we demonstrated a potential method for rapidly assigning function to different PULs, the ultimate goal is to expand the assigned function to the vast number of PULs. This can be achieved by further utilizing the scouting enzyme and cognate HTCS sensor domain specificity to a larger number of PULs, thus generating a polysaccharide degradation map of Bacteroidetes species. This broader understanding, combined with metabolic profiles of Bacteroidetes grown in different polysaccharides, may be used to shed light on the microbial metabolic network. This potential metabolic mapping in the gastrointestinal tract can be combined with metagenomic analysis of gut comunities to efficiently manipulate the Bacteroidetes community in the gut, potentially affecting the entire microbiome.

REFERENCES

Anders, S., D. J. McCarthy, Y. Chen, M. Okoniewski, G. K. Smyth, W. Huber and M. D. Robinson (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.,* **8**(9): 1765-1786.

Anders, S., P. T. Pyl and W. Huber (2015). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics,* **31**(2): 166-169.

Anderson, K. L. and A. A. Salyers (1989). Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer-membrane starch-binding sites and periplasmic starch-degrading enzymes. *J. Bacteriol.,* **171**(6): 3192-3198.

Andrade, C. M. C., W. B. Aguiar and G. Antranikian (2001). Physiological aspects involved in production of xylanolytic enzymes by deep-sea hyperthermophilic archaeon Pyrodictium abyssi. *Appl. Biochem. Biotechnol*., **91**(1-9): 655-669.

Arpaia, N., C. Campbell, X. Fan, S. Dikiy, J. van der Veeken, P. deRoos, H. Liu, J. R. Cross, K. Pfeffer, P. J. Coffer and A. Y. Rudensky (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*, **504**(7480): 451-455.

Baughn, A. D. and M. H. Malamy (2004). The strict anaerobe *Bacteroides fragilis* grows in and benefits from nanomolar concentrations of oxygen. *Nature*, **427**(6973): 441-444.

Bergman, E. N. (1990). Energy contributions of volatile fatty-acids from the gastrointestinal-tract in various species. *Physiol. Rev.* **70**(2): 567-590.

Bhattacharya, T., T. S. Ghosh and S. S. Mande (2015). Global profiling of carbohydrate active enzymes in human gut microbiome. *PLoS One,* **10**(11): e0142038.

Bjursell, M. K., E. C. Martens and J. I. Gordon (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J. Biol. Chem.,* **281**(47): 36269-36279.

Bloemen, J. G., K. Venema, M. C. van de Poll, S. W. Olde Damink, W. A. Buurman and C. H. Dejong (2009). Short chain fatty acids exchange across the gut and liver in humans measured at surgery. *Clin. Nutr.,* **28**(6): 657-661.

Bolger, A. M., M. Lohse and B. Usadel (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114-2120.

Bonawitz, N. D. and C. Chapple (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* **44**: 337-363.

Cann, I. K., K. Komori, H. Toh, S. Kanai and Y. Ishino (1998). A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc. Nat. Acad. Sci. U. S. A.,* **95**(24): 14250-14255.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.,* **37**(Database issue): D233-238.

Cerdeno-Tarraga, A. M., S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell and J. Parkhill (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, **307**(5714): 1463-1465.

Chen, F., W. Wu and Y. Cong (2016). Short chain fatty acids regulation of neutrophil production of IL-10, *Am. Assoc. Immnol*.

Chesson, A., A.H. Gordon, and J.A. Lomax, (1985). Methylation analysis of mesophyll, epidermis, and fiber cell-walls isolated from the leaves of perennial and italian. *Carbohydr. Res.*, **141(1)**: 137-147.

Cho, K. H. and A. A. Salyers (2001). Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.,* **183**(24): 7224-7230.

Clausen, M. R. and P. Mortensen (1995). Kinetic studies on colonocyte metabolism of short chain fatty acids and glucose in ulcerative colitis. *Gut*, **37**(5): 684-689.

Cobb, B. A., Q. Wang, A. O. Tzianabos and D. L. Kasper (2004). Polysaccharide processing and presentation by the MHCII pathway. *Cell,* **117**(5): 677-687.

Cox, L. M. and M. J. Blaser (2013). Pathways in microbe-induced obesity. *Cell Metab.,* **17**(6): 883-894.

Cummings, J. and G. Macfarlane (1997). Role of intestinal bacteria in nutrient metabolism. *Clin. Nutr.,* **16**(1): 3-11.

D'Elia, J. N. and A. A. Salyers (1996). Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J. Bacteriol.,* **178**(24): 7173-7179.

D'Elia, J. N. and A. A. Salyers (1996). Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.,* **178**(24): 7180-7186.

Dasgupta, S., D. Erturk-Hasdemir, J. Ochoa-Reparaz, H. C. Reinecker and D. L. Kasper (2014). Plasmacytoid dendritic cells mediate anti-inflammatory responses to a gut

commensal molecule via both innate and adaptive mechanisms. *Cell Host Microbe,* **15**(4): 413-423.

Davies G. J., W. K., Henrissat B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem*. *J.,* : 557-559.

De Filippo, C., D. Cavalieri, M. Di Paola, M. Ramazzotti, J. B. Poullet, S. Massart, S. Collini, G. Pieraccini and P. Lionetti (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci. U. S. A.,* **107**(33): 14691-14696.

de Vries, R. P., P. A. vanKuyk, H. C. M. Kester and J. Visser (2002). The *Aspergillus niger faeB* gene encodes a second feruloyl esterase involved in pectin and xylan degradation and is specifically induced in the presence of aromatic compounds. *Biochem. J.,* **363**(2): 377-386.

Despres, J., E. Forano, P. Lepercq, S. Comtet-Marre, G. Jubelin, C. J. Yeoman, M. E. B. Miller, C. J. Fields, N. Terrapon and C. Bourvellec (2016). "Unraveling the pectinolytic function of Bacteroides xylanisolvens using a RNA-seq approach and mutagenesis." BMC genomics **17**(1): 147.

Dodd, D. and I. K. Cann (2009). Enzymatic deconstruction of xylan for biofuel production. *Glob. Change Biol. Bioenergy,* **1**(1): 2-17.

Dodd, D., M. H. Spitzer, W. Van Treuren, B. D. Merrill, A. J. Hryckowian, S. K. Higginbottom, A. Le, T. M. Cowan, G. P. Nolan and M. A. Fischbach (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature*, **551**(7682): 648.

Donath, M. Y., E. Dalmas, N. S. Sauter and M. Boni-Schnetzler (2013). Inflammation in obesity and diabetes: islet dysfunction and therapeutic opportunity. *Cell Metab*., **17**(6): 860-872.

Dong, G. C., C. Y. Kuan, S. Subramaniam, J. Y. Zhao, S. Sivasubramaniam, H. Y. Chang and F. H. Lin (2015). A potent inhibition of oxidative stress induced gene expression in neural cells by sustained ferulic acid release from chitosan based hydrogel. *Mater. Sci. Eng. C. Mater. Biol. Appl.,* **49**: 691-699.

Donohoe, D. R., L. B. Collins, A. Wali, R. Bigler, W. Sun and S. J. Bultman (2012). The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol. Cell*, **48**(4): 612-626.

Dornez, E., K. Gebruers, S. Cuyvers, J. A. Delcour and C. M. Courtin (2007). Impact of wheat flour-associated endoxylanases on arabinoxylan in dough after mixing and resting. *J. Agric. Food. Chem*., **55**(17): 7149-7155.

Duncan, S. H., G. Holtrop, G. E. Lobley, A. G. Calder, C. S. Stewart and H. J. Flint (2004). Contribution of acetate to butyrate formation by human faecal bacteria. *Br. J. Nutr*., **91**(6): 915-923.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson and D. A. Relman (2005). Diversity of the human intestinal microbial flora. *Science*, **308**(5728): 1635-1638.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*., **32**(5): 1792-1797.

El Kaoutari, A., F. Armougom, J. I. Gordon, D. Raoult and B. Henrissat (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Rev. Microbiol.*, **11**(7): 497.

Faith, J. J., J. L. Guruge, M. Charbonneau, S. Subramanian, H. Seedorf, A. L. Goodman, J. C. Clemente, R. Knight, A. C. Heath, R. L. Leibel, M. Rosenbaum and J. I. Gordon (2013). The long-term stability of the human gut microbiota. *Science* **341**(6141): 1237439.

Ferreira, C. M., A. T. Vieira, M. A. R. Vinolo, F. A. Oliveira, R. Curi and F. d. S. Martins (2014). The central role of the gut microbiota in chronic inflammatory diseases. *J. Immunol. Res.,* **2014**.

Flegal, K. M., D. Kruszon-Moran, M. D. Carroll, C. D. Fryar and C. L. Ogden (2016). Trends in Obesity Among Adults in the United States, 2005 to 2014. *JAMA,* **315**(21): 2284-2291.

Flint, H. J., K. P. Scott, S. H. Duncan, P. Louis and E. Forano (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes,* **3**(4): 289-306.

Fujimoto, Z., S. Kaneko, A. Kuno, H. Kobayashi, I. Kusakabe and H. Mizuno (2004). Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86. *J. Biol. Chem.,* **279**(10): 9606-9614.

Fukuda, S., H. Toh, K. Hase, K. Oshima, Y. Nakanishi, K. Yoshimura, T. Tobe, J. M. Clarke, D. L. Topping, T. Suzuki, T. D. Taylor, K. Itoh, J. Kikuchi, H. Morita, M. Hattori and H. Ohno (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature,* **469**(7331): 543-547.

Furusawa, Y., Y. Obata, S. Fukuda, T. A. Endo, G. Nakato, D. Takahashi, Y. Nakanishi, C. Uetake, K. Kato, T. Kato, M. Takahashi, N. N. Fukuda, S. Murakami, E. Miyauchi, S.

Hino, K. Atarashi, S. Onawa, Y. Fujimura, T. Lockett, J. M. Clarke, D. L. Topping, M. Tomita, S. Hori, O. Ohara, T. Morita, H. Koseki, J. Kikuchi, K. Honda, K. Hase and H. Ohno (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature,* **504**(7480): 446-450.

Gani, A., W. Sm and M. Fa (2012). Whole-grain cereal bioactive compounds and their health benefits: a review. *J. Food Process. Technol.*, **03**(03).

Gaudier, E., A. Jarry, H. Blottiere, P. De Coppet, M. Buisine, J. Aubert, C. Laboisse, C. Cherbut and C. Hoebler (2004). Butyrate specifically modulates MUC gene expression in intestinal epithelial goblet cells deprived of glucose. *Am. J. Physiol. Gastrointest. Liver Physiol.,* **287**(6): G1168-G1174.

Gill, S. C. and P. H. Von Hippel (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.,* **182**(2): 319-326.

Gill, S. R., M. Pop, R. T. Deboy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett and K. E. Nelson (2006). Metagenomic analysis of the human distal gut microbiome. *Science,* **312**(5778): 1355-1359.

Glenwright, A. J., K. R. Pothula, S. P. Bhamidimarri, D. S. Chorev, A. Baslé, S. J. Firbank, H. Zheng, C. V. Robinson, M. Winterhalter and U. Kleinekathöfer (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*, **541**(7637): 407.

Grondin, J. M., K. Tamura, G. Dejean, D. W. Abbott and H. Brumer (2017). Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol*., **199**(15).

Guarner, F. and J.-R. Malagelada (2003). Gut flora in health and disease. *Lancet,* **361**(9356): 512-519.

Guinane, C. M. and P. D. Cotter (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap*. *Adv*. *Gastroenterol*., **6**(4): 295-308.

Henrissat, B. and A. Bairoch (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem*. *J*., **316 (Pt 2)**: 695-696.

Hoch, J. A. and T. J. Silhavy (1995). Two-component signal transduction, *ASM press* Washington, DC:.

Hold, G. L., S. E. Pryde, V. J. Russell, E. Furrie and H. J. Flint (2002). Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. *FEMS Microbiol. Ecol.,* **39**(1): 33-39.

Hooper, L. V., D. R. Littman and A. J. Macpherson (2012). Interactions between the microbiota and the immune system. *Science,* **336**(6086): 1268-1273.

Hooper, L. V., T. Midtvedt and J. I. Gordon (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.,* **22**: 283-307.

Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature,* **486**(7402): 207-214.

Izydorczyk, M. S. and C. G. Biliaderis (1995). Cereal arabinoxylans: advances in structure and physicochemical properties. *Carbohydr. Polym.,* **28**(1): 33-48.

Jin Son, M., Rico, C. W., S. Hyun Nam and M. Young Kang (2010). Influence of oryzanol and ferulic acid on the lipid metabolism and antioxidative status in high fat-fed mice. *J. Clin. Biochem. Nutr.,* **46**(2): 150-156.

Johnson, K., J. Fontana and C. MacKenzie (1988). Measurement of acetylxylan esterase in Streptomyces. *Methods Enzymol.*, Elsevier. **160:** 551-560.

Jonsson, A. L. and F. Backhed (2017). Role of gut microbiota in atherosclerosis. *Nat. Rev. Cardiol.,* **14**(2): 79-87.

Jordan, D. B. and K. Wagschal (2010). Properties and applications of microbial beta-D-xylosidases featuring the catalytically efficient enzyme from *Selenomonas ruminantium*. *Appl. Microbiol. Biotechnol.,* **86**(6): 1647-1658.

Kabel, M. A., C. J. Yeoman, Y. Han, D. Dodd, C. A. Abbas, J. A. de Bont, M. Morrison, I. K. Cann and R. I. Mackie (2011). Biochemical characterization and relative expression levels of multiple carbohydrate esterases of the xylanolytic rumen bacterium Prevotella ruminicola 23 grown on an ester-enriched substrate. *Appl. Environ. Microbiol.,* **77**(16): 5671-5681.

Kalina, U., N. Koyama, T. Hosoda, H. Nuernberger, K. Sato, D. Hoelzer, F. Herweck, T. Manigold, M. V. Singer and S. Rossol (2002). Enhanced production of IL-18 in butyrate-treated intestinal epithelium by stimulation of the proximal promoter region. *Eur. J. Immunol.*, **32**(9): 2635-2643.

Kanauchi, O., T. Suga, M. Tochihara, T. Hibi, M. Naganuma, T. Homma, H. Asakura, H. Nakano, K. Takahama and Y. Fujiyama (2002). Treatment of ulcerative colitis by feeding with germinated barley foodstuff: first report of a multicenter open control trial. *J. Gastroenterol.,* **37**(14): 67-72.

Kaneko, S., H. Ichinose, Z. Fujimoto, A. Kuno, K. Yura, M. Go, H. Mizuno, I. Kusakabe and H. Kobayashi (2004). Structure and function of a family 10 beta-xylanase chimera of *Streptomyces olivaceoviridis* E-86 FXYN and *Cellulomonas fimi* Cex. *J. Biol. Chem.,* **279**(25): 26619-26626.

Karlsson, F., V. Tremaroli, J. Nielsen and F. Backhed (2013). Assessing the human gut microbiota in metabolic diseases. *Diabetes,* **62**(10): 3341-3349.

Katoh, K., K. Misawa, K. i. Kuma and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.,* **30**(14): 3059-3066.

Khan, M. T., M. Nieuwdorp and F. Backhed (2014). Microbial modulation of insulin sensitivity. *Cell Metab.,* **20**(5): 753-760.

Kim, M., Y. Qie, J. Park and C. H. Kim (2016). Gut Microbial Metabolites Fuel Host Antibody Responses. *Cell Host Microbe,* **20**(2): 202-214.

Koeth, R. A., Z. Wang, B. S. Levison, J. A. Buffa, E. Org, B. T. Sheehy, E. B. Britt, X. Fu, Y. Wu, L. Li, J. D. Smith, J. A. DiDonato, J. Chen, H. Li, G. D. Wu, J. D. Lewis, M. Warrier, J. M. Brown, R. M. Krauss, W. H. Tang, F. D. Bushman, A. J. Lusis and S. L. Hazen (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.,* **19**(5): 576-585.

Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.,* **28**(4): 416-436.

Kotarski, S. F. and A. A. Salyers (1984). Isolation and characterization of outer membranes of *Bacteroides thetaiotaomicron* grown on different carbohydrates. *J. Bacteriol*., **158**(1): 102-109.

Kraut, J. (1977). Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.,* **46**(1): 331-358.

Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V. K. Sharma, T. P. Srivastava, T. D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D. S. Ehrlich,

K. Itoh, T. Takagi, Y. Sakaki, T. Hayashi and M. Hattori (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.,* **14**(4): 169-181.

Kuwahara, T., A. Yamashita, H. Hirakawa, H. Nakayama, H. Toh, N. Okada, S. Kuhara, M. Hattori, T. Hayashi and Y. Ohnishi (2004). Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. U. S. A.,* **101**(41): 14919-14924.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* **227**(5259): 680.

Langmead, B. and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods,* **9**(4): 357-359.

Larsbrink, J., T. E. Rogers, G. R. Hemsworth, L. S. McKee, A. S. Tauzin, O. Spadiut, S. Klinter, N. A. Pudlo, K. Urs, N. M. Koropatkin, A. L. Creagh, C. A. Haynes, A. G. Kelly, S. N. Cederholm, G. J. Davies, E. C. Martens and H. Brumer (2014). A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature,* **506**(7489): 498-502.

Le Poul, E., C. Loison, S. Struyf, J. Y. Springael, V. Lannoy, M. E. Decobecq, S. Brezillon, V. Dupriez, G. Vassart, J. Van Damme, M. Parmentier and M. Detheux (2003). Functional characterization of human receptors for short chain fatty acids and their role in polymorphonuclear cell activation. *J. Biol. Chem.,* **278**(28): 25481-25489.

Lee, C.-C., C.-C. Wang, H.-M. Huang, C.-L. Lin, S.-J. Leu and Y.-L. Lee (2015). Ferulic acid induces Th1 responses by modulating the function of dendritic cells and ameliorates

Th2-mediated allergic airway inflammation in mice. *Evid. Based Complement. Alternat. Med.,* **2015**.

Lee, S. M., G. P. Donaldson, Z. Mikulski, S. Boyajian, K. Ley and S. K. Mazmanian (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature,* **501**(7467): 426-429.

Letunic, I. and P. Bork (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.,* **44**(W1): W242-W245.

Levy, M., C. A. Thaiss and E. Elinav (2016). Metabolites: messengers between the microbiota and the immune system. Genes & development **30**(14): 1589-1597.

Ley, R. E., F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight and J. I. Gordon (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U. S. A.,* **102**(31): 11070-11075.

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel and R. Knight (2008). Evolution of mammals and their gut microbes. *Science,* **320**(5883): 1647-1651.

Ley, R. E., P. J. Turnbaugh, S. Klein and J. I. Gordon (2006). Microbial ecology: human gut microbes associated with obesity. *Nature,* **444**(7122): 1022.

Liming Xia, P. C. (1999). Cellulase production by solid state fermentation on lignocellulosic waste from the xylose industry. Adv Biochem Eng/Biotechnol **65**: 69-92.

Lloyd-Price, J., G. Abu-Ali and C. Huttenhower (2016). The healthy human microbiome. *Genome Med.,* **8**(1): 51.

Löser, C., A. Eisel, D. Harms and U. Fölsch (1999). Dietary polyamines are essential luminal growth factors for small intestinal and colonic mucosal growth and development. *Gut*, **44**(1): 12-16.

Louis, P., G. L. Hold and H. J. Flint (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.,* **12**(10): 661.

Lowe, E. C., A. Baslé, M. Czjzek, S. J. Firbank and D. N. Bolam (2012). A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system. *Proc. Nat. Acad. Sci. U. S. A.,* **109**(19): 7298-7303.

Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson and R. Knight (2012). Diversity, stability and resilience of the human gut microbiota. *Nature,* **489**(7415): 220-230.

Lynch, J. B. and J. L. Sonnenburg (2012). "Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system." Molecular microbiology **85**(3): 478-491.

Macfarlane, G. T., S. Hay, S. Macfarlane and G. R. Gibson (1990). Effect of different carbohydrates on growth, polysaccharidase and glycosidase production by *Bacteroides ovatus*, in batch and continuous culture. *J. Appl. Bacteriol.,* **68**(2): 179-187.

Macia, L., J. Tan, A. T. Vieira, K. Leach, D. Stanley, S. Luong, M. Maruya, C. Ian McKenzie, A. Hijikata, C. Wong, L. Binge, A. N. Thorburn, N. Chevalier, C. Ang, E. Marino, R. Robert, S. Offermanns, M. M. Teixeira, R. J. Moore, R. A. Flavell, S. Fagarasan and C. R. Mackay (2015). Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nat. Commun.,* **6**: 6734.

Mackie, R. I. and I. Cann (2018). Let them eat fruit. *Nat. Microbiol.,* **3**(2): 127.

Manetti, R., P. Parronchi, M. G. Giudizi, M. Piccinni, E. Maggi, G. Trinchieri and S. Romagnani (1993). Natural killer cell stimulatory factor (interleukin 12 [IL-12]) induces T helper type 1 (Th1)-specific immune responses and inhibits the development of IL-4-producing Th cells. *J. Experiment. Med.,* **177**(4): 1199-1204.

Martens, E. C., H. C. Chiang and J. I. Gordon (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe, **4**(5): 447-457.

Martens, E. C., E. C. Lowe, H. Chiang, N. A. Pudlo, M. Wu, N. P. McNulty, D. W. Abbott, B. Henrissat, H. J. Gilbert, D. N. Bolam and J. I. Gordon (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.,* **9**(12): e1001221.

Maslowski, K. M., A. T. Vieira, A. Ng, J. Kranich, F. Sierro, D. Yu, H. C. Schilter, M. S. Rolph, F. Mackay, D. Artis, R. J. Xavier, M. M. Teixeira and C. R. Mackay (2009). Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature,* **461**(7268): 1282-1286.

Matsumoto, M., R. Kibe, T. Ooga, Y. Aiba, S. Kurihara, E. Sawaki, Y. Koga and Y. Benno (2012). Impact of intestinal microbiota on intestinal luminal metabolome. *Sci. Rep*., **2**: 233.

Mazmanian, S. K. and D. L. Kasper (2006). The love-hate relationship between bacterial polysaccharides and the host immune system. *Nat. Rev. Immunol.,* **6**(11): 849-858.

Mazmanian, S. K., J. L. Round and D. L. Kasper (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature,* **453**(7195): 620-625.

McNulty, N. P., M. Wu, A. R. Erickson, C. Pan, B. K. Erickson, E. C. Martens, N. A. Pudlo, B. D. Muegge, B. Henrissat and R. L. Hettich (2013). Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11**(8): e1001637.

Michael, A. J. (2016). Biosynthesis of polyamines and polyamine-containing molecules. *Biochem. J.,* **473**(15): 2315-2329.

Miller-Fleming, L., V. Olin-Sandoval, K. Campbell and M. Ralser (2015). Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *J. Mol. Biol.,* **427**(21): 3389-3406.

Moore, K. J., F. J. Sheedy and E. A. Fisher (2013). Macrophages in atherosclerosis: a dynamic balance. *Nat*. *Rev*. *Immunol*., **13**(10): 709-721.

Moore, W. E. C., Holdeman, L. V. (1974). Human fecal flora: the normal flora of 20 japanese-hawaiian. *Appl. Microbiol.,* **27**: 961-979.

Morrison, D. J. and T. Preston (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes,* **7**(3): 189-200.

Ndeh, D., A. Rogowski, A. Cartmell, A. S. Luis, A. Basle, J. Gray, I. Venditto, J. Briggs, X. Zhang, A. Labourel, N. Terrapon, F. Buffetto, S. Nepogodiev, Y. Xiao, R. A. Field, Y. Zhu, M. A. O'Neil, B. R. Urbanowicz, W. S. York, G. J. Davies, D. W. Abbott, M. C. Ralet, E. C. Martens, B. Henrissat and H. J. Gilbert (2017). Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature,* **544**(7648): 65-70.

Nugent, S. G. (2001). Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. *Gut,* **48**(4): 571-577.

Ohnishi, M., T. Matuo, T. Tsuno, A. Hosoda, E. Nomura, H. Taniguchi, H. Sasaki and H. Morishita (2004). Antioxidant activity and hypoglycemic effect of ferulic acid in STZ-induced diabetic mice and KK-Aymice. *BioFactors,* **21**(1-4): 315-319.

Pandey, M. P. and C. S. Kim (2011). Lignin depolymerization and conversion: a review of thermochemical methods. *Chem. Eng. Technol.,* **34**(1): 29-41.

Pavlostathis, S. G., T. L. Miller and M. J. Wolin (1988). Fermentation of insoluble cellulose by continuous cultures of *Ruminococcus albus*. *Appl. Environ. Microbiol.,* **54**(11): 2655-2659.

Pell, G., L. Szabo, S. J. Charnock, H. Xie, T. M. Gloster, G. J. Davies and H. J. Gilbert (2004). Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J. Biol. Chem.,* **279**(12): 11777-11788.

Pell, G., E. J. Taylor, T. M. Gloster, J. P. Turkenburg, C. M. Fontes, L. M. Ferreira, T. Nagy, S. J. Clark, G. J. Davies and H. J. Gilbert (2004). The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. *J. Biol. Chem.,* **279**(10): 9597-9605.

Perez-Cano, F. J., A. Gonzalez-Castro, C. Castellote, A. Franch and M. Castell (2010). Influence of breast milk polyamines on suckling rat immune system maturation. *Dev. Comp. Immunol.,* **34**(2): 210-218.

Polizeli, M. L., A. C. Rizzatti, R. Monti, H. F. Terenzi, J. A. Jorge and D. S. Amorim (2005). Xylanases from fungi: properties and industrial applications. *Appl. Microbiol. Biotechnol.,* **67**(5): 577-591.

Price, M. N., P. S. Dehal and A. P. Arkin (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One,* **5**(3): e9490.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, H. I. T. C. Meta, P. Bork, S. D. Ehrlich and J. Wang (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* **464**(7285): 59-65.

Qin, J., Y. Li, Z. Cai, S. Li, J. Zhu, F. Zhang, S. Liang, W. Zhang, Y. Guan, D. Shen, Y. Peng, D. Zhang, Z. Jie, W. Wu, Y. Qin, W. Xue, J. Li, L. Han, D. Lu, P. Wu, Y. Dai, X. Sun, Z. Li, A. Tang, S. Zhong, X. Li, W. Chen, R. Xu, M. Wang, Q. Feng, M. Gong, J. Yu, Y. Zhang, M. Zhang, T. Hansen, G. Sanchez, J. Raes, G. Falony, S. Okuda, M. Almeida, E. LeChatelier, P. Renault, N. Pons, J. M. Batto, Z. Zhang, H. Chen, R. Yang, W. Zheng, S. Li, H. Yang, J. Wang, S. D. Ehrlich, R. Nielsen, O. Pedersen, K. Kristiansen and J. Wang (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature,* **490**(7418): 55-60.

Ramon, D., P. vd Veen and J. Visser (1993). Arabinan degrading enzymes from *Aspergillus nidulans*: induction and purification. *FEMS Microbiol. Lett.,* **113**(1): 15-22.

Reeves, A. R., J. N. DElia, J. Frias and A. A. Salyers (1996). A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J. Bacteriol.,* **178**(3): 823-830.

Richard Wolfenden, X. L., and Gregory Young (1998). Spontaneous Hydrolysis of Glycosides. *J. Am. Chem. Soc.,* **120**: 6814-6815.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1): 139-140.

Rocha, E. R., T. Selby, J. P. Coleman and C. J. Smith (1996). Oxidative stress response in an anaerobe, Bacteroides fragilis: a role for catalase in protection against hydrogen peroxide. *J. Bacteriol.,* **178**(23): 6895-6903.

Rogowski, A., J. A. Briggs, J. C. Mortimer, T. Tryfona, N. Terrapon, E. C. Lowe, A. Basle, C. Morland, A. M. Day, H. Zheng, T. E. Rogers, P. Thompson, A. R. Hawkins, M. P. Yadav, B. Henrissat, E. C. Martens, P. Dupree, H. J. Gilbert and D. N. Bolam (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun*., **6**: 7481.

Russell, W. R., S. H. Duncan, L. Scobbie, G. Duncan, L. Cantlay, A. G. Calder, S. E. Anderson and H. J. Flint (2013). Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol. Nutr. Food Res.,* **57**(3): 523-535.

Saha, B. C. (2000). α-l-Arabinofuranosidases. Biotechnology Advances **18**(5): 403-423.

Sallusto, F. and A. Lanzavecchia (1994). Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. *J. Experiment. Med.,* **179**(4): 1109-1118.

Salyers, A. A. (1984). *Bacteroides* of the human lower intestinal tract. *Annu. Rev. Microbiol.,* **38**: 293-313.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.,* **31**(1): 107-133.

Schwalm, N. D., G. E. Townsend and E. A. Groisman (2017). Prioritization of polysaccharide utilization and control of regulator activation in *Bacteroides thetaiotaomicron*. *Mol. Microbiol.,* **104**(1): 32-45.

Serrano, A., C. Palacios, G. Roy, C. Cespón, M. a. L. Villar, M. Nocito and P. González-Porqué (1998). Derivatives of gallic acid induce apoptosis in tumoral cell lines and inhibit lymphocyte proliferation. *Arch. Biochem. Biophys.,* **350**(1): 49-54.

Shipman, J. A., J. E. Berleman and A. A. Salyers (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.,* **182**(19): 5365-5372.

Sina, C., O. Gavrilova, M. Forster, A. Till, S. Derer, F. Hildebrand, B. Raabe, A. Chalaris, J. Scheller, A. Rehmann, A. Franke, S. Ott, R. Hasler, S. Nikolaus, U. R. Folsch, S. Rose-John, H. P. Jiang, J. Li, S. Schreiber and P. Rosenstiel (2009). G protein-coupled receptor 43 is essential for neutrophil recruitment during intestinal inflammation. *J. Immunol.,* **183**(11): 7514-7522.

Smith, C. E. and K. L. Tucker (2011). Health benefits of cereal fibre: a review of clinical trials. *Nutr. Res. Rev.,* **24**(1): 118-131.

Smith, P. M., M. R. Howitt, N. Panikov, M. Michaud, C. A. Gallini, M. Bohlooly-y, J. N. Glickman and W. S. Garrett (2013). The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science,*: 1237242.

Sokol, H., P. Seksik, J. Furet, O. Firmesse, I. Nion-Larmurier, L. Beaugerie, J. Cosnes, G. Corthier, P. Marteau and J. Doré (2009). Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm. Bowel Dis.,* **15**(8): 1183-1189.

Soliman, M. L., K. L. Puig, C. K. Combs and T. A. Rosenberger (2012). Acetate reduces microglia inflammatory signaling in vitro. *J. Neurochem.,* **123**(4): 555-567.

Stalikas, C. D. (2007). Extraction, separation, and detection methods for phenolic acids and flavonoids. *J.* Separat. *Science,* **30**(18): 3268-3295.

Sun, J., C. Tian, S. Diamond and N. L. Glass (2012). Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. *Eukaryot. Cell,* **11**(4): 482-493.

Sun, Y. and J. Cheng (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.,* **83**(1): 1-11.

Tegtmeier, D., C. L. Thompson, C. Schauer and A. Brune (2015). Oxygen affects gut bacterial colonization and metabolic activities in a gnotobiotic cockroach model. *Appl. Environ. Microbiol.,* **82**(4): 1080-1089.

Terrapon, N., V. Lombard, H. J. Gilbert and B. Henrissat (2015). Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics,* **31**(5): 647-655.

Thorvaldsdottir, H., J. T. Robinson and J. P. Mesirov (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.*, **14**(2): 178-192.

Thygesen, A., A. B. Thomsen, A. S. Schmidt, H. Jørgensen, B. K. Ahring and L. Olsson (2003). Production of cellulose and hemicellulose-degrading enzymes by filamentous fungi cultivated on wet-oxidised wheat straw. *Enzyme Microb. Technol.,* **32**(5): 606-615.

Topping, D. L. and P. M. Clifton (2001). Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.,* **81**(3): 1031-1064.

Turnbaugh, P. J., F. Backhed, L. Fulton and J. I. Gordon (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe*, **3**(4): 213-223.

Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight and J. I. Gordon (2009). A core gut microbiome in obese and lean twins. *Nature,* **457**(7228): 480-484.

Vardakou, M., C. Dumon, J. W. Murray, P. Christakopoulos, D. P. Weiner, N. Juge, R. J. Lewis, H. J. Gilbert and J. E. Flint (2008). Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases. *J. Mol. Biol.,* **375**(5): 1293-1305.

Vardakou, M., J. Flint, P. Christakopoulos, R. J. Lewis, H. J. Gilbert and J. W. Murray (2005). A family 10 *Thermoascus aurantiacus* xylanase utilizes arabinose decorations of xylan as significant substrate specificity determinants. *J. Mol. Biol.,* **352**(5): 1060-1067.

Vinolo, M. A., H. G. Rodrigues, E. Hatanaka, F. T. Sato, S. C. Sampaio and R. Curi (2011). Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *J. Nutr. Biochem.,* **22**(9): 849-855.

Vrieze, A., E. Van Nood, F. Holleman, J. Salojarvi, R. S. Kootte, J. F. Bartelsman, G. M. Dallinga-Thie, M. T. Ackermans, M. J. Serlie, R. Oozeer, M. Derrien, A. Druesne, J. E. Van Hylckama Vlieg, V. W. Bloks, A. K. Groen, H. G. Heilig, E. G. Zoetendal, E. S. Stroes, W. M. de Vos, J. B. Hoekstra and M. Nieuwdorp (2012). Transfer of intestinal microbiota

from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterol.,* **143**(4): 913-916 e917.

Wang, K., G. V. Pereira, J. J. V. Cavalcante, M. Zhang, R. Mackie and I. Cann (2016). *Bacteroides intestinalis* DSM 17393, a member of the human colonic microbiome, upregulates multiple endoxylanases during growth on xylan. *Sci. Rep.,* **6**: 34360.

Wang, Z., E. Klipfell, B. J. Bennett, R. Koeth, B. S. Levison, B. Dugar, A. E. Feldstein, E. B. Britt, X. Fu, Y. M. Chung, Y. Wu, P. Schauer, J. D. Smith, H. Allayee, W. H. Tang, J. A. DiDonato, A. J. Lusis and S. L. Hazen (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature,* **472**(7341): 57-63.

Wefers, D., J. J. Cavalcante, R. R. Schendel, J. Deveryshetty, K. Wang, Z. Wawrzak, R. I. Mackie, N. M. Koropatkin and I. Cann (2017). Biochemical and Structural Analyses of Two Cryptic Esterases in Bacteroides intestinalis and their Synergistic Activities with Cognate Xylanases. *J. Mol. Biol.,* **429**(16): 2509-2527.

West, C. E., M. C. Jenmalm and S. L. Prescott (2015). The gut microbiota and its role in the development of allergic disease: a wider perspective. *Clin. Exp. Allergy,* **45**(1): 43-53.

Wexler, H. M. (2007). *Bacteroides*: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.,* **20**(4): 593-621.

Whitehead, T. R. (1995). Nucleotide sequence of xylan-inducible xylanase and xylosidase/arabinosidase gene from *Bacteroides ovatus*. *Biochim. Biophys. Acta,* **1244**: 239-241.

Wikoff, W. R., A. T. Anfora, J. Liu, P. G. Schultz, S. A. Lesley, E. C. Peters and G. Siuzdak (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.,* **106**(10): 3698-3703.

Willemsen, L., M. Koetsier, S. Van Deventer and E. Van Tol (2003). Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E1 and E2 production by intestinal myofibroblasts. *Gut,* **52**(10): 1442-1447.

Wrzosek, L., S. Miquel, M.-L. Noordine, S. Bouet, M. J. Chevalier-Curt, V. Robert, C. Philippe, C. Bridonneau, C. Cherbuy and C. Robbe-Masselot (2013). *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol.,* **11**(1): 61.

Wu, G. D., J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S. A. Keilbaugh, M. Bewtra, D. Knights, W. A. Walters and R. Knight (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science,* **334**(6052): 105-108.

Wu, M., N. P. McNulty, D. A. Rodionov, M. S. Khoroshkin, N. W. Griffin, J. Cheng, P. Latreille, R. A. Kerstetter, N. Terrapon, B. Henrissat, A. L. Osterman and J. I. Gordon (2015). Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science,* **350**(6256): aac5992.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper and J. I. Gordon (2003). A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science,* **299**(5615): 2074-2076.

Xu, J., M. A. Mahowald, R. E. Ley, C. A. Lozupone, M. Hamady, E. C. Martens, B. Henrissat, P. M. Coutinho, P. Minx, P. Latreille, H. Cordum, A. Van Brunt, K. Kim, R. S. Fulton, L. A. Fulton, S. W. Clifton, R. K. Wilson, R. D. Knight and J. I. Gordon (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.,* **5**(7): e156.

Yasuko, K., N. Tomohiro, M. Sei-Itsu, L. Ai-Na, F. Yasuo and T. Takashi (1984). Caffeic acid is a selective inhibitor for leukotriene biosynthesis. *Biochim. Biophys. Acta Lipids Lipid Metab*., **792**(1): 92-97.

Yoshida, S., R. I. Mackie and I. K. Cann (2010). Biochemical and domain analyses of FSUAxe6B, a modular acetyl xylan esterase, identify a unique carbohydrate binding module in *Fibrobacter succinogenes* S85. *J. Bacteriol.,* **192**(2): 483-493.

Zhang, F., W. Luo, Y. Shi, Z. Fan and G. Ji (2012). Should we standardize the 1,700-year-old fecal microbiota transplantation? The American *J. Gastroenterol.,* **107**(11): 1755.

Zhang, M., L. V. Borovikova, H. Wang, C. Metz and K. J. Tracey (1999). Spermine inhibition of monocyte activation and inflammation. *Mol. Med.,* **5**(9): 595.

Zhang, M., J. R. Chekan, D. Dodd, P. Y. Hong, L. Radlinski, V. Revindran, S. K. Nair, R. I. Mackie and I. Cann (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci. U. S. A.,* **111**(35): E3708-3717.

Zhu, W., J. C. Gregory, E. Org, J. A. Buffa, N. Gupta, Z. Wang, L. Li, X. Fu, Y. Wu, M. Mehrabian, R. B. Sartor, T. M. McIntyre, R. L. Silverstein, W. H. W. Tang, J. A. DiDonato, J. M. Brown, A. J. Lusis and S. L. Hazen (2016). Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. *Cell*, **165**(1): 111-124.

APPENDIX A: MEDIA COMPOSITION

Supplemental Table A.1. Brain heart infusion supplemented (BHIS) medium

Ingredient	Concentration (L^{-1})	Concentration (M)
Brain heart infusion	37 g	
Hemin chloride	1.2 mg	1.9 µM
L-Cysteine Hydrochloride	0.63 g	4 mM
Sodium Carbonate	4 g	38 mM

Supplemental Table A.2. Composition of defined medium used for the culture of *Bacteroides* spp.

Ingredient	Concentration (g/L)	Concentration (M)
Carbon Source[a,b,c,d,e, f, g, h]	5 g	0.5% wt/vol
Ammonium Sulfate[a]	1.32 g	10 mM
Sodium Chloride	0.875 g	15 mM
Dipotassium Hydrogen Phosphate	0.871 g	5 mM
Monopotassium Hydrogen Phosphate	0.68 g	5 mM
Sodium Carbonate	0.4 g	3.8 mM
Sodium Sulfide	0.5 g	6.4 mM
Magnesium (II) Chloride Heptahydrate	20 mg	0.1 mM
Calcium (II) Chloride Dihydrate	8 mg	54 µM
Iron (II) Sulphate Heptahydrate	0.4 mg	1.4 µM
Hemin Chloride	1.2 mg	1.9 µM
Vitamin K_3 (Menadione)	1 mg	5.8 µM
Vitamin B_{12}	10 µg	7.3 nM

[a] Used xylose:arabinose mixture (62:48) as sole carbon source.
[b] Used insoluble wheat arabinoxylan as sole carbon source.
[c] Used soluble wheat arabinoxylan as sole carbon source.
[d] Used wheat bran as sole carbon source.
[e] Used destarched wheat bran as sole carbon source.
[f] Used arabinose as sole carbon source.
[g] Used arabinan as sole carbon source.
[h] Used ferulated oligosaccharides as sole carbon source.

APPENDIX B: TRANSCRIPTOMICS ANALYSIS PIPELINE

- **Compile and concatenate ribosomal RNA sequences is a fasta file**

Combine ribosomal RNA in a single file to remove ribosomal reads that have not been removed by Ribozero library construction.

- **Quality control – FastQC**

Load raw reads from sequencing center with FastQC software. This software will demonstrate the quality of the reads. By assessing the quality of the reads, any region bellow quality score 30 can be trimmed to not affect the alignment.

- **Remove rRNA from raw reads**

Adding bowtie2 to current directory

```
export PATH=$PATH:/path/to/bowtie2
```

Building rRNA into bowtie2

```
bowtie2-build -f "/path/to/rRNA.txt" RiboRiddance
```

Remove rRNA reads from quality checked data

```
bowtie2 -x RiboRiddance -U "/path/to/QCreads.fastq" --un "trimmed_QCreads.fq" -S rRNA.sam
```

- **Align sequences – bowtie2**

Building genome into bowtie2

```
bowtie2-build -f "/path/to/genome.fasta" Genome
```

Align trimmed reads to genome

```
bowtie2 -x Genome -U "/path/to/trimmed_QCreads.fastq" -S trimmed_QCreads.sam
```

Convert sam reads to bam reads

```
export PATH=$PATH:"/path/to/samtools
samtools view -b trimmed_QCreads.sam > trimmed_QCreads.bam
```

- **Sort and Index files**

Sort reads in bam file

```
samtools sort -m 5000000000 trimmed_QCreads.bam -o trimmed_QCreads_sorted.bam
```

Index reads of sorted bam file

```
samtools index trimmed_QCreads_sorted.bam trimmed_QCreads.bai
```

- **Visualize read coverage map using Integrative Genome Viewer**

IGV requires both .bam and .bai files to be in the same folder. Furthermore, the genome file (gff + fasta file) need to be combined into a .genome file generated by IGV.

- **Count aligned reads**

Count reads from sam file

```
htseq-count -s no -I locus_tag -t gene trimmed_QCreads.sam Genome.gff3 > trimmed_QCreads.count
```

- **TPM/RPKM analysis**

Calculate TPM and RPKM in R/Rstudio

Make a file containing gene and length (locus tag and kb) and save as .csv

Make a file containing counts and save as .csv

Both files should have matching rows

Add function to R for TPM

```
rpkm <- function(counts, lengths) {
rate <- counts / lengths
rate / sum(counts) * 1e6
}
```

Add function to R for RPKM

```
tpm <- function(counts, lengths) {
```

```
rate <- counts / lengths
rate / sum(rate) * 1e6
}
```

Add gene table to R

```
genes = read.csv("/path/to/genetable.csv")
```

#Add count table to R

```
counts = read.csv("/path/to/countable.csv")
```

Calculated TPM

```
tpms <- apply(counts, 2, function(x) tpm(x, genes$Length))
```

#Calculate RPKM

```
rpkms <- apply(counts, 2,function(x) rpkm(x, genes$Length))
```

Write TPM values

```
write.csv(tpms, file = "TPM.csv")
```

Write RPKM values

```
write.csv(rpkms, file = "RPMK.csv")
```

- **Differential expression comparison**

Load required libraries

```
library("limma")
library("edgeR")
```

Clear the environment

```
rm(list=ls())
```

Preparation <- simple design (2 conditions)

```
samples =read.csv("matrix.csv", stringsAsFactors=F)
samples$countf = paste(samples$SampleName, "count", sep=".")
counts = readDGE(samples$countf)$counts
```

Filter weakly expressed

```
noint = rownames(counts) %in%
  c("no_feature", "ambiguous", "too_low_aQual", "not_aligned",
"alignment_not_unique")
cpms = cpm(counts)
keep = rowSums(cpms >1) >=2 & !noint
counts = counts[keep,]

# At this point, inspect the count table

colnames(counts) = samples$SampleName
head(counts[,order(samples$Condition)], 5)

# Start of analysis

d = DGEList(counts=counts, group = samples$Condition)
e = calcNormFactors(d)
plotMDS(e,labels=samples$shortname,
col=c("darkgreen","blue")[factor(samples$Condition)])
f = estimateCommonDisp(e)
g = estimateTagwiseDisp(f)
plotMeanVar(g, show.tagwise.vars = TRUE, NBline = TRUE)
plotBCV(g)
h = exactTest(g, pair = c("CTL", "TRT"))

# TopTags function is utilized to present a tabular summary of the DE stats

tt= topTags(h, n=nrow(g))
head(tt$table)

# Expect output including logFC, logCPM, LR, PValue and FDR

# Normalized depth-adjusted reads per million

nc = cpm(g, normalized.lib.sizes = TRUE)
rn = rownames(tt$table)
head(nc[rn,order(samples$Condition)],5)

# Create a graphical summary

deg = rn[tt$table$FDR < .05]
plotSmear(g, de.tags=deg)

# Save output file

write.csv(tt$table, file="toptags_edgeR.csv")
```

APPENDIX C: BACTEROIDES SPECIES GENES UP-REGULATED DURING GROWTH ON INSOLUBLE WHEAT ARABINOXYLAN COMPARED TO THE MONOSACCHARIDE (XYLOSE:ARABINOSE)

Supplemental Table C.1. Genes of *B. intestinalis* with significant transcriptional changes during growth in insoluble wheat arabinoxylan compared to the monosaccharides mixture.

Locus tag	Fold change	Predicted function	P-value
BACINT_01657	228.7	hypothetical protein	1.70E-39
BACINT_01039	118.5	CE1	6.25E-06
BACINT_01035	111.9	GH43|CBM13	4.36E-05
BACINT_01041	99.0	GH43	6.74E-06
BACINT_01037	98.5	SusD	0.000122254
BACINT_01040	92.7	hypothetical protein	8.70E-06
BACINT_01034	90.0	hypothetical protein	6.44E-05
BACINT_01038	85.1	CE6|CE1	7.79E-06
BACINT_01033	72.2	CE1	0.000137462
BACINT_01042	71.0	GH3	3.37E-06
BACINT_01036	68.3	SusC	5.85E-05
BACINT_01043	66.7	GH43|CBM6	3.71E-07
BACINT_01856	39.9	RNA Binding	2.23E-16
BACINT_00014	31.9	RNA Binding	1.10E-22
BACINT_04204	31.7	hypothetical protein	5.34E-25
BACINT_03234	26.3	hypothetical protein	1.83E-19
BACINT_04603	21.1	hypothetical protein	7.97E-14
BACINT_04205	20.6	GH67	1.74E-18
BACINT_04562	20.6	hypothetical protein	8.55E-18
BACINT_04203	20.1	GH43	9.90E-29
BACINT_00588	19.8		7.07E-19
BACINT_02097	18.6	hypothetical protein	3.97E-17
BACINT_02234	18.2		7.65E-24
BACINT_01094	16.8	CE1|CE1	2.41E-17

Supplemental Table C.1. (Cont.)

BACINT_04202	15.3	GH10|GH43	2.41E-14
BACINT_00039	15.1	hypothetical protein	3.86E-21
BACINT_01596	14.0	S15P	3.02E-12
BACINT_01617	13.8	Hbs	5.26E-15
BACINT_03683	13.8	hypothetical protein	2.55E-14
BACINT_00431	13.1	hypothetical protein	8.62E-14
BACINT_02960	12.9	SusD	8.19E-30
BACINT_04201	12.9	MFS	1.95E-15
BACINT_01386	12.7	hypothetical protein	4.88E-11
BACINT_03685	12.6	hypothetical protein	1.20E-14
BACINT_02771	12.6	hypothetical protein	3.54E-19
BACINT_00927	12.6	GH8	9.70E-22
BACINT_00772	12.0	hypothetical protein	8.36E-25
BACINT_01986	11.8	hypothetical protein	8.48E-17
BACINT_00620	11.8		1.24E-17
BACINT_00890	11.7	S20P	9.62E-11
BACINT_00583	11.5	L19P	1.93E-10
BACINT_02776	11.2	hypothetical protein	5.08E-24
BACINT_00310	11.2	hypothetical protein	1.40E-08
BACINT_01814	10.7	hypothetical protein	9.21E-32
BACINT_03376	10.6	L-alanine dehydrogenase	1.18E-12
BACINT_02386	10.6	hypothetical protein	2.41E-12
BACINT_00302	10.4	hypothetical protein	1.21E-30

Supplemental Table C.2. Genes of *B. cellulosilyticus* with significant transcriptional changes during growth on insoluble wheat arabinoxylan compared to the monosaccharides mixture.

Locus tag	Fold change	Predicted function	P-value
BACCELL_00298	664.3	hypothetical protein	4.24E-11
BACCELL_02764	660.2	hypothetical protein	3.91E-18
BACCELL_03406	233.0	SusC	3.46E-106
BACCELL_02430	146.0	SusC	2.73E-120
BACCELL_01332	129.6	SusC	4.37E-132
BACCELL_03411	124.3	hypothetical protein	6.81E-53
BACCELL_03530	119.3	hypothetical protein	2.30E-102
BACCELL_03407	119.1	SusD	8.94E-111
BACCELL_02062	107.4	CE1|CE1	1.46E-74
BACCELL_03531	98.3	hypothetical protein	1.95E-109
BACCELL_03409	92.0	SusD	4.63E-99
BACCELL_01031	86.2	Protease inhibitor	2.81E-109
BACCELL_03410	85.9	hypothetical protein	2.11E-100
BACCELL_03408	80.1	SusC	3.03E-74
BACCELL_02143	79.8	GH43	2.17E-95
BACCELL_02261	79.0	GH8	1.26E-123
BACCELL_02152	77.6	GH43|CBM13	4.00E-90
BACCELL_02431	74.6	SusD	3.77E-139
BACCELL_02432	74.2	hypothetical protein	2.86E-110
BACCELL_01333	73.2	SusD	1.94E-98
BACCELL_03438	71.4	GH35	5.43E-117
BACCELL_02144	71.4	CE1	3.97E-95
BACCELL_02153	71.3	hypothetical protein	1.59E-80
BACCELL_02141	71.1	GH43|CBM6	5.17E-96
BACCELL_03437	70.8	hypothetical protein	8.59E-101
BACCELL_03663	70.5	hypothetical protein	3.52E-51

Supplemental Table C.2. (Cont.)

BACCELL_01212	69.8	hypothetical protein	3.64E-139
BACCELL_02154	69.4	CE1	8.46E-92
BACCELL_02147	69.0	GH9	1.70E-82
BACCELL_02145	68.8	CE1	4.88E-89
BACCELL_02061	67.8	GH43	2.50E-115
BACCELL_05446	66.4	SusC	3.46E-100
BACCELL_02148	64.8	SusD	4.50E-73
BACCELL_02146	63.6	hypothetical protein	9.14E-91
BACCELL_01458	60.7	hypothetical protein	1.62E-25
BACCELL_05210	60.6	hypothetical protein	8.37E-74
BACCELL_05444	59.5	hypothetical protein	8.93E-126
BACCELL_01116	55.2	GH10/CE1	2.71E-75
BACCELL_05443	54.8	hypothetical protein	4.34E-121
BACCELL_01655	54.5	SusC	8.49E-108
BACCELL_03436	53.7	Sugar transporter	1.19E-105
BACCELL_05700	51.8	Bacterial DNA-binding protein.	1.82E-27
BACCELL_03766	50.9	Protein of unknown function	2.33E-34
BACCELL_03441	49.9	hypothetical protein	6.78E-113
BACCELL_03635	47.6	hypothetical protein	8.91E-21
BACCELL_00461	47.4	hypothetical protein	5.60E-31
BACCELL_03671	46.5	hypothetical protein	2.08E-47

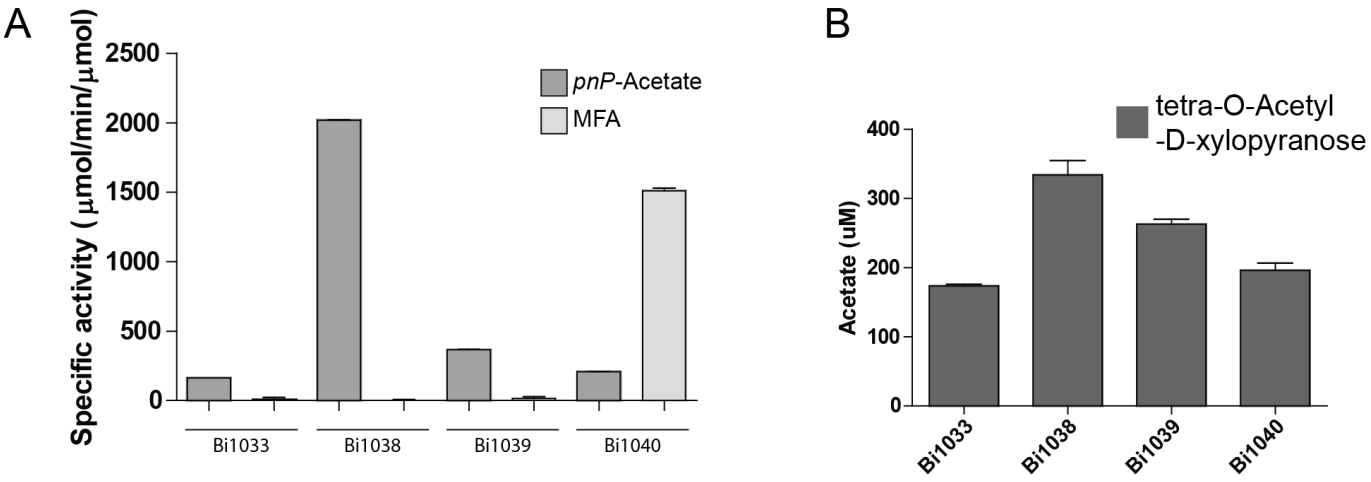
Supplemental Table C.3. Genes of *B. oleiciplenus* with significant transcriptional changes during growth on insoluble wheat arabinoxylan compared to the monosaccharides mixture.

Locus tag	Fold change	Putative function	P-value
HMPREF9447_02595	77.4	SusD	1.28E-34
HMPREF9447_02596	71.4	hypothetical protein	6.90E-27
HMPREF9447_02528	59.8	GH43	4.06E-29
HMPREF9447_02593	50.9	GH2	5.50E-27
HMPREF9447_02594	48.7	SusC	2.08E-25
HMPREF9447_02527	45.4	GH3	1.05E-32
HMPREF9447_02529	44.9	GH9	5.59E-40
HMPREF9447_02818	44.7	hypothetical protein	2.02E-47
HMPREF9447_02530	38.2	GH43|CBM13	6.47E-35
HMPREF9447_02531	38.2	hypothetical protein	8.37E-41
HMPREF9447_01474	35.2	SusC	3.20E-11
HMPREF9447_03068	32.9	elongation factor	1.97E-38
HMPREF9447_05303	32.6	SusC	3.46E-29
HMPREF9447_04842	32.1	hypothetical protein	6.94E-42
HMPREF9447_02526	30.9	GH43|CBM6	3.32E-38
HMPREF9447_02534	30.6	SusC	6.97E-37
HMPREF9447_04880	29.4	hypothetical protein	1.35E-23
HMPREF9447_01633	28.8	hypothetical protein	2.64E-36
HMPREF9447_04828	28.7	hypothetical protein	4.94E-40
HMPREF9447_02325	28.4	hypothetical protein	2.80E-14
HMPREF9447_02597	28.4	GH51	5.98E-24
HMPREF9447_00592	27.5	hypothetical protein	2.04E-39
HMPREF9447_03111	25.7	Anti-sigma	2.96E-28
HMPREF9447_02532	23.8	CE1	2.39E-38
HMPREF9447_02492	23.7	hypothetical protein	2.19E-32
HMPREF9447_01458	23.7	MFS	1.19E-29
HMPREF9447_01734	23.0	hypothetical protein	4.57E-34

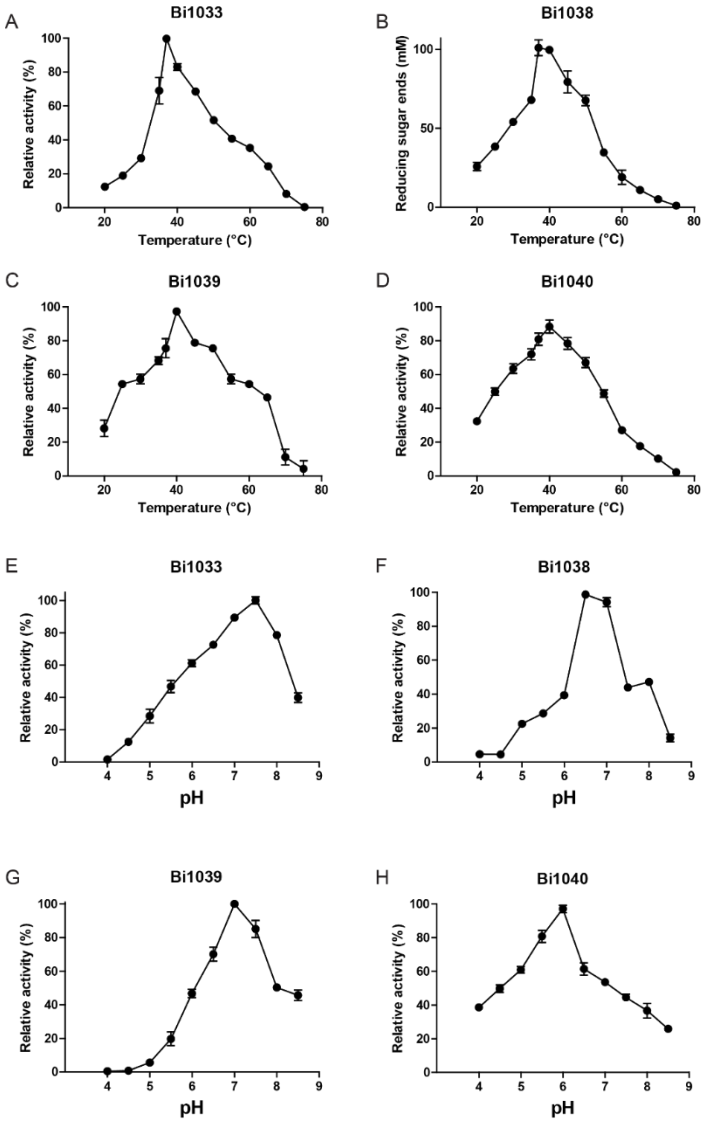
Supplemental Table C.3. (Cont.)

HMPREF9447_02533	22.9	SusD	3.94E-38
HMPREF9447_04843	22.1	hypothetical protein	4.76E-35
HMPREF9447_04836	20.2	hypothetical protein	1.13E-26
HMPREF9447_01462	19.6	GH67	1.00E-26
HMPREF9447_02224	19.4	SusD	2.58E-13
HMPREF9447_01453	18.9	SusD	7.69E-29
HMPREF9447_01452	18.3	SusC	1.54E-27
HMPREF9447_00748	18.2	GH18	5.04E-27
HMPREF9447_02660	17.3	hypothetical protein	2.10E-20
HMPREF9447_01456	16.7	GH35	2.14E-23
HMPREF9447_04831	16.4	hypothetical protein	1.82E-24
HMPREF9447_04890	16.1	hypothetical protein	1.24E-19
HMPREF9447_04699	16.1	hypothetical protein	3.89E-23
HMPREF9447_05150	16.0	Anti-sigma	2.06E-30
HMPREF9447_01455	15.6	GH10|CBM22|CBM22	4.05E-24
HMPREF9447_02223	15.6	SusC	5.95E-17
HMPREF9447_00921	15.1	hypothetical protein	3.76E-29
HMPREF9447_02490	15.1	hypothetical protein	4.58E-28
HMPREF9447_00593	14.8	hypothetical protein	1.08E-29
HMPREF9447_04835	14.4	hypothetical protein	1.42E-17
HMPREF9447_01829	14.3	GH43	6.94E-08

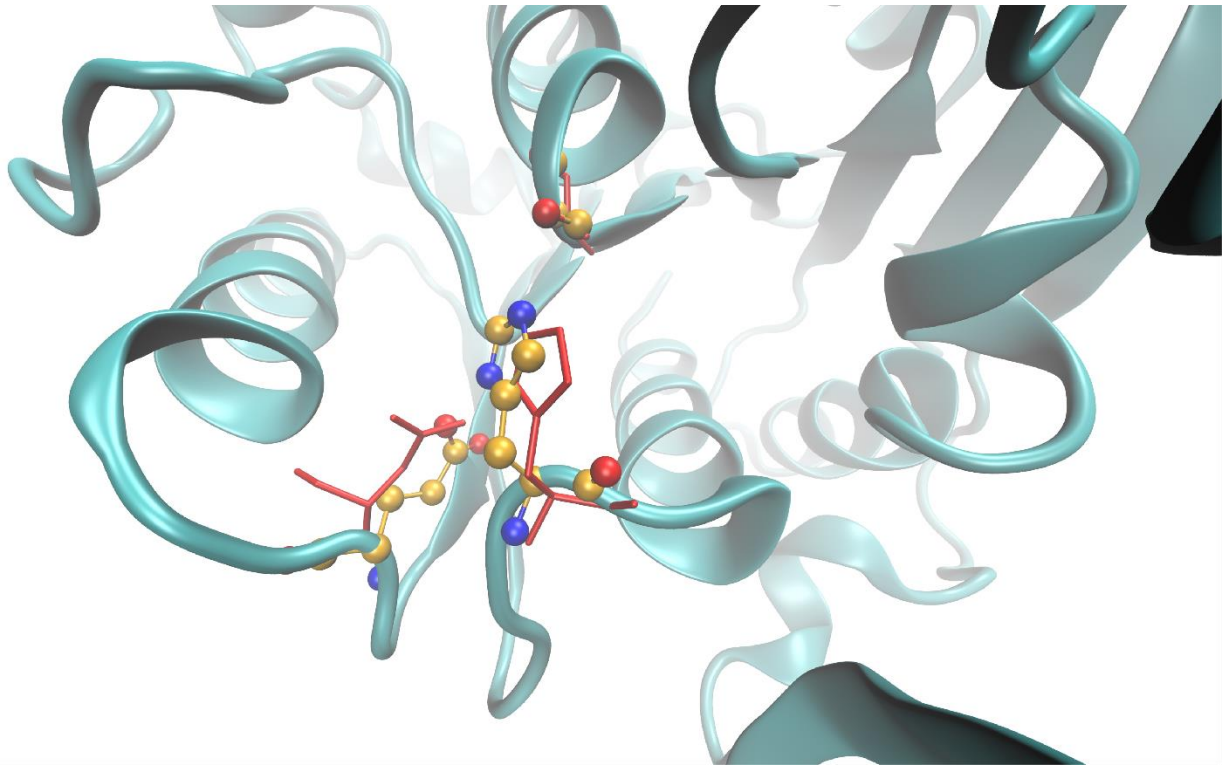
APPENDIX D: SUPPLEMENTAL FIGURES



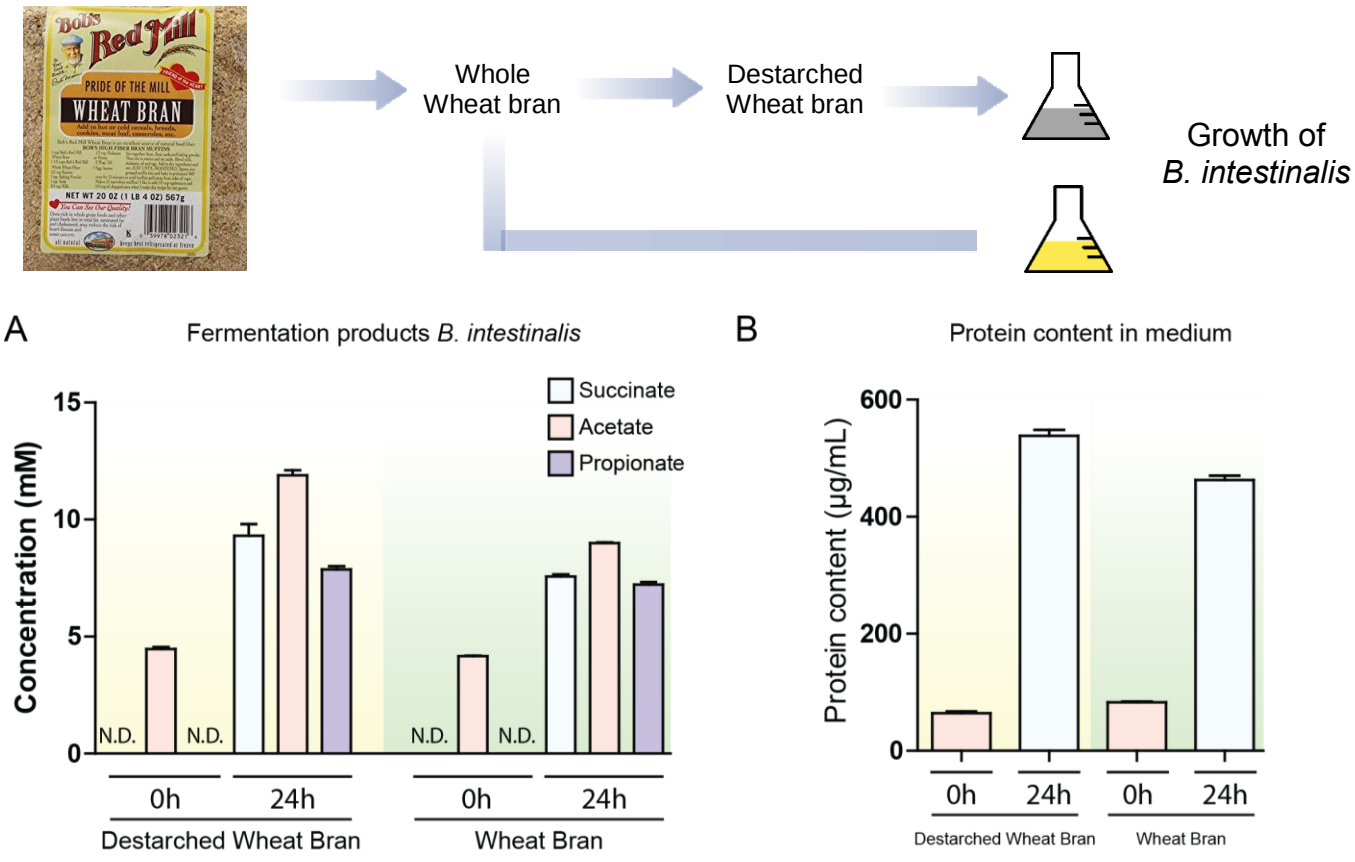
Supplemental Figure D.1 - **Hydrolytic activity of esterases towards synthetic substrates.** (A) Specific activity of the putative esterases towards methyl-ferulate and *p*NP-acetate. (B) Substrate specificity of putative esterases towards synthetic tetra-O-acetyl-D-xylopyranose.



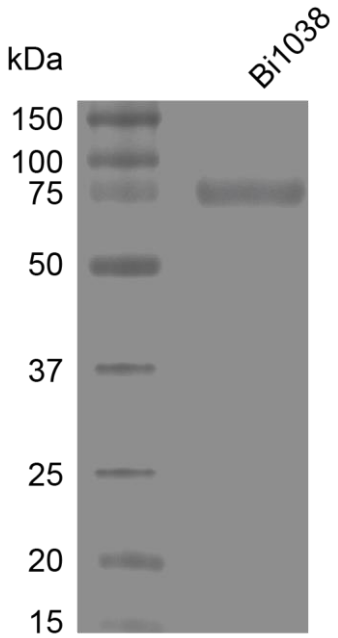
Supplemental Figure D.2 - **Biochemical characterization of putative esterases in the esterase-enriched cluster.** Enzymatic activity reported relative to the optimal pH (A) and temperature (B). Error bars represent the standard deviation of three replicates.



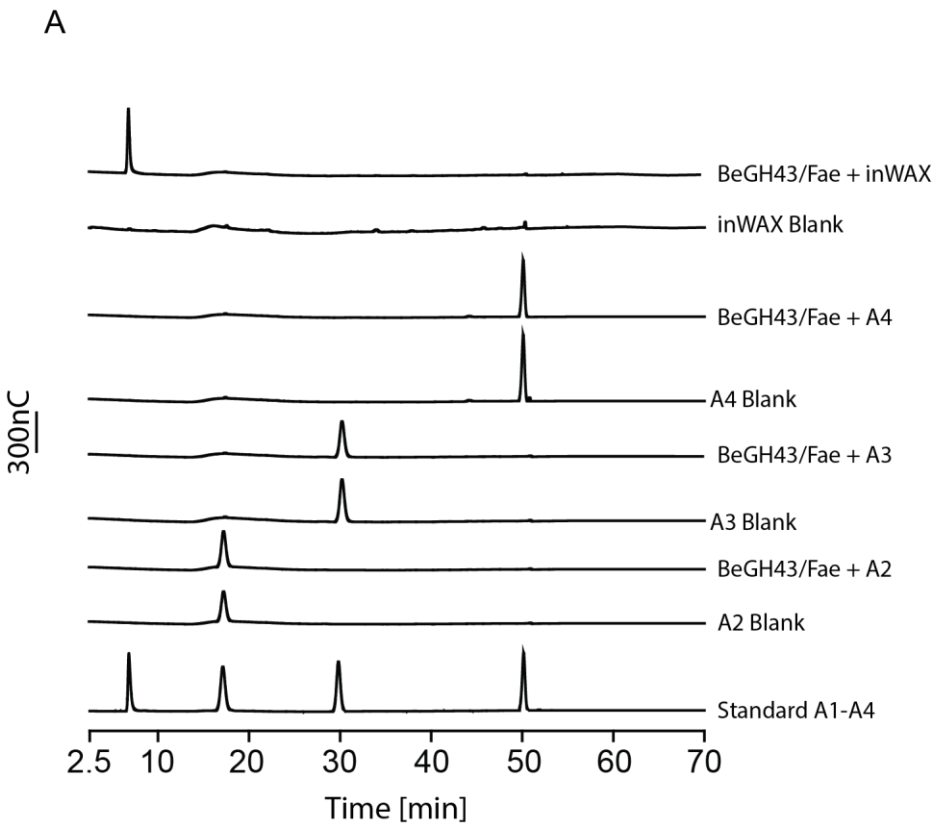
Supplemental Figure D.3 - **Superimposition of two acetyl xylan esterases.** The superimposition of the acetyl xylan esterase encoded by BACINT_01039 (teal) and catalytic triad (red) from C. thermocellum esterase demonstrate conservation. Glutamate, serine and histidine amino acids from Bi1039 are depicted with its specific atoms (yellow: carbon, blue: nitrogen, and red: oxygen).



Supplemental Figure D.4 - ***B. intestinalis*** **is able to grow in minimal medium containing either wheat bran or destarched wheat bran as the sole carbon source.** (A) *B. intestinalis* growth in complex substrates demonstrate short-chain fatty acid production and (B) increase in the protein content in the medium, showing growth in these substrates.



Supplemental Figure D.5 - **Serine Protease inhibitor, Benzamidine HCl, prevents cleavage of Bi1038 construct during expression and purification in the E. coli heterologous expression system.** SDS gel depicts the purification of the full construct by adding 1mM of Benzamidine HCl to buffers used during cell resuspension, lysis, and purification steps.



Supplemental Figure D.6. **GH43 domain of BeGH43/Fae demonstrate arabinoxylan dependent arabinofuranosidase activity.** (A) HPLC chromatogram shows activity of BeGH43/Fae releasing arabinose (A1) from insoluble wheat arabinoxylan (inWAX), while demonstrating no activity towards arabinooligosaccharides (A2-A4).